Singular integrals with flag kernels
on homogeneous groups, I

Alexander Nagel, Fulvio Ricci, Elias M. Stein and Stephen Wainger

Abstract. Let $\mathcal{K}$ be a flag kernel on a homogeneous nilpotent Lie group G. We prove that operators T of the form $T(f) = f * \mathcal{K}$ form an algebra under composition, and that such operators are bounded on $L^p(G)$ for $1 < p < \infty$.

1. Introduction

This is the first of two papers dealing with singular integral operators with flag kernels on homogeneous nilpotent groups. Our goal is to show that these operators, along with appropriate sub-collections, form algebras under composition, and that the operators in question are bounded on L^p.

Operators of this kind arose initially when studying compositions of sub-elliptic operators on the Heisenberg group (such as the sub-Laplacian $\mathcal{L}$ and $\Box_b$) with elliptic-type operators. In particular in [7] one saw that operators of the form $m(\mathcal{L}, iT)$ (where m is a "Marcinkiewicz multiplier") are singular integrals with flag kernels and satisfy L^p estimates. The theory was extended in [9] to encompass general flag kernels in the Euclidean space $\mathbb{R}^N$, and the resulting operators arising via abelian convolution. In addition, aspects of the CR theory for quadratic manifolds could be studied via such operators on various 2-step groups. More recently, flag kernels have been studied in [13] and [2]. In view of this, and because of their potential further application, it is desirable to extend the above results in [9] to the setting of homogeneous groups of higher step. To achieve this goal requires however that we substantially recast the approach and techniques used previously, since these were essentially limited to the 2-step case.

Our main results are two-fold. Suppose G is a homogeneous nilpotent group and $\mathcal{K}$ denotes a distribution on G which is a flag kernel (the requisite definitions are given below in Definition 2.3).

Theorem A. *The operators T of the form $T(f) = f * \mathcal{K}$ form an algebra under composition.*

Theorem B. *The above operators are bounded on $L^p(G)$ for $1 < p < \infty$.*

Given the complexity of the material, in this introductory section we provide the reader with an outline of the main ideas that enter in the proofs of the above theorems. Moreover, in order to simplify the presentation we will often not state matters in the most general setting and sometimes describe the situation at hand a little imprecisely.

1.1. Flag kernels

We start with a direct sum decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$, with $\sum_{j=1}^n a_j = N$, and we write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, with $\mathbf{x}_m \in \mathbb{R}^{a_m}$. We also fix a one-parameter family of dilations δ_r on $\mathbb{R}^N$, given by $\delta_r(\mathbf{x}) = (r^{d_1}\mathbf{x}_1, \ldots, r^{d_n}\mathbf{x}_n)$, with positive exponents $d_1 < d_2 < \cdots < d_n$.[1] We denote by $Q_k = d_k a_k$ the homogeneous dimension of $\mathbb{R}^{a_k}$. We also define the partial "norms" $N_k(\mathbf{x}) = |\mathbf{x}_k|_e^{1/d_k}$, where $|\mathbf{x}_k|_e$ is the standard Euclidean norm on $\mathbb{R}^{a_k}$.

In this setting, a flag kernel $\mathcal{K}$ is a distribution on $\mathbb{R}^N$ which is given by integration against a C^∞ function $K(\mathbf{x})$ away from $\mathbf{x}_1 = 0$ and which satisfies two types of conditions. The first are the differential inequalities for $\mathbf{x}_1 \neq 0$:

$$(1.1) \qquad |\partial_{\mathbf{x}}^\alpha K(\mathbf{x})| \leq C_\alpha \prod_{k=1}^n \left(N_1(\mathbf{x}) + N_2(\mathbf{x}) + \cdots + N_k(\mathbf{x})\right)^{-Q_k - d_k \alpha_k}$$

with $\alpha = (\alpha_1, \ldots, \alpha_n)$. The second are the cancellation conditions. These are most easily expressed recursively. Let $\langle \mathcal{K}, \varphi \rangle$ denote the action of the distribution $\mathcal{K}$ on a test function φ. At the beginning of the recursion there is the following condition, in many ways typical of the others:

$$(1.2) \qquad \sup_R |\langle \mathcal{K}, \varphi_R \rangle| < \infty$$

where $R = (R_1, R_2, \ldots, R_n)$, $\varphi_R(x) = \varphi(R_1^{d_1}x_1, R_2^{d_2}x_2, \ldots, R_n^{d_n}x_n)$, and φ is an arbitrary C^∞ function which is supported in the unit ball. More generally, one requires that the action of $\mathcal{K}$ on a test function in some subset of variables $\{\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}\}$ produces a flag kernel in the remaining variables $\{\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_\alpha}\}$. The precise formulation of these conditions is given in Section 2 and Definition 2.3 below.

1.2. Dyadic decomposition

A main tool used in studying flag kernels is their dyadic decomposition into sums of "bump functions". This proceeds as follows. Let $I = (i_1, i_2, \ldots, i_n)$ denote any indexing set of integers that satisfies.

$$(1.3) \qquad i_1 \leq i_2 \leq \cdots \leq i_{n-1} \leq i_n.$$

---

[1]One can also allow non-isotropic dilations on each subspace $\mathbb{R}^{a_l}$. See Section 2 below.

Also let $\{\varphi^I\}$ be a family of C^∞ functions supported in the unit ball that are uniformly bounded in the $C^{(m)}$ norm for each m. Set

$$[\varphi^I]_I(\mathbf{x}) = 2^{-i_1 Q_1 - i_2 Q_2 - \cdots - i_n Q_n} \; \varphi^I(2^{-d_1 i_1}\mathbf{x}_1, \ldots, 2^{-d_n i_n}\mathbf{x}_n),$$

so that the $[\varphi^I]_I$ are L^1-normalized. We say that the φ^I satisfy the "strong cancellation" condition if for each k with $1 \le k \le n$,

$$(1.4) \qquad \int_{\mathbb{R}^{a_k}} \varphi^I(\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_n)\, d\mathbf{x}_k \equiv 0$$

when all the inequalities (1.3) for I are strict. In the case that there are some equalities in (1.3), say $i_{\ell-1} < i_\ell = i_{\ell+1} = \cdots = i_k < i_{k+1}$, then only cancellation in the collection of corresponding variables is required:

$$(1.4') \qquad \int_{\mathbb{R}^{a_\ell} \oplus \cdots \oplus \mathbb{R}^{a_k}} \varphi^I(\mathbf{x}_1, \ldots, \mathbf{x}_\ell, \ldots, \mathbf{x}_k, \ldots, \mathbf{x}_n)\, d\mathbf{x}_\ell \cdots d\mathbf{x}_k \equiv 0.$$

The first result needed is that any sum

$$(1.5) \qquad \qquad \qquad \sum_I [\varphi^I]_I$$

made up of such bump functions, with the cancellation conditions (1.4) and (1.4'), converges in the sense of distributions to a flag kernel, and conversely, any flag kernel $\mathcal{K}$ can be written in this way (of course, not uniquely).

There are two parts to this result (which in effect is stated but not proved completely in [9]). The first is that the sum in (1.5) is indeed a flag kernel. To see this, one can use the estimate in Proposition 10.1 given in Appendix II below; one also notes from this that even without the cancellation conditions (1.4) and (1.4'), the sum (1.5) satisfies the differential inequalities (1.1). The converse part requires Theorem 6.1 below, and the observation that the parts of the sum (1.5) contributed by I's where there may be equality in (1.3) give flag distributions corresponding to various "coarser" flags.

However, what will be key in what follows is that the strong cancellation conditions (1.4) or (1.4') can be weakened, and still lead to the same conclusion. While these "weak" cancellation conditions are somewhat complicated to state (see Definition 5.5 below), they are easily illustrated in the special 2-step case. Here we have the decomposition $\mathbb{R}^n = \mathbb{R}^{a_1} \oplus \mathbb{R}^{a_2}$, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$. The cancellation condition for the second variable is as before: $\int \varphi^I(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_2 \equiv 0$. For $\mathbf{x}_1$ the weak cancellation condition takes the form

$$(1.6) \qquad \int_{\mathbb{R}^{a_1}} \varphi^I(\mathbf{x}_1, \mathbf{x}_2)\, d\mathbf{x}_1 = 2^{-\epsilon(i_2 - i_1)}\eta^I(\mathbf{x}_2),$$

for some $\epsilon > 0$, with $I = (i_1, i_2)$ and η^I an L^1 normalized bump in the $\mathbf{x}_2$ variable.

In this context, the main conclusion (Theorem 6.8) is that the sum (1.5) is still a flag kernel if the weak-cancellation conditions are assumed instead of (1.4)

and $(1.4')$, and the functions $\{\varphi^I\}$ are allowed to belong to the Schwartz class instead of being compactly supported. In understanding Definition 5.5, one should keep in mind that conditions like (1.4), which involve vanishing of integrals, are equivalent with expressions of the φ^I as the sums of appropriate derivatives. (This is established in Lemma 5.1).

1.3. Other properties of flag kernels

Along with the results about decompositions of flag kernels, there are a number of other properties of these distributions that are worth mentioning and are discussed in Section 6. First, the class of flag kernels is invariant under changes of variables compatible with the structure of the flags. We have in mind transformations $\mathbf{x} \mapsto \mathbf{y} = F(\mathbf{x})$, with $\mathbf{y}_k = \mathbf{x}_k + P_k(\mathbf{x})$, and P_k a homogeneous polynomial of $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$, of the same homogeneity as $\mathbf{x}_k$. The fact that $\mathcal{K} \circ F$ satisfies the same differential inequalities (1.1) as $\mathcal{K}$ is nearly obvious, but the requisite cancellation conditions (such as (1.2)) are subtler and involve the weak cancellation of the bump functions. (See Theorem 6.15.)

A second fact is that the cancellations required in the definition of a flag kernel can be relaxed. For example, assuming that the differential inequalities (1.1) hold, then the less restrictive version of (1.2) requires that the supremum is taken only over those R for which $R_1 \geq R_2 \geq \cdots \geq R_n > 0$. The formulation and proof of the sufficiency of these restricted conditions is in Theorem 6.13.

Finally we should point out that at the basis of many of our arguments is an earlier characterization in [9] of flag kernels in terms of their Fourier transforms: these are bounded multipliers that satisfy the dual differential inequalities given in Definition 6.3.

1.4. Graded groups and compositions of flag kernels

Up to this point our discussion of flag kernels has focused on their definition as distributions on the Euclidean space $\mathbb{R}^N$. We now consider convolutions with flag kernels on graded nilpotent Lie groups G whose underlying space is $\mathbb{R}^N$. The choice of an appropriate coordinate system on the group G, and its multiplication structure, induces a decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$ and allows us to find exponents $d_1 < d_2 < \cdots < d_n$ as above so that the dilations $\delta_r(\mathbf{x}) = (r^{d_1}\mathbf{x}_1, \ldots, r^{d_n}\mathbf{x}_n)$, with $r > 0$, are automorphisms of G.

The proof of Theorem A reduces to the statement that if $\mathcal{K}_1$ and $\mathcal{K}_2$ are a pair of flag kernels, then $\mathcal{K}_1 * \mathcal{K}_2$ is a sum of flag kernels, where the convolution is taken with respect to G. Note that when G is the abelian group $\mathbb{R}^N$, the result follows immediately from the characterization of flag kernels in terms of their Fourier transforms, cited earlier, and in fact the convolution of two flag kernels is a single flag kernel. In the non-commutative case the proof is not as simple and proceeds as follows. First write $\mathcal{K}_1 = \sum_I [\varphi^I]_I$, $\mathcal{K}_2 = \sum_J [\psi^J]_J$ in terms of decompositions with bump functions with strong cancellation. Now, formally,

$$(1.7) \qquad \mathcal{K}_1 * \mathcal{K}_2 = \sum_I \sum_J [\varphi^I]_I * [\psi^J]_J .$$

We look first at an individual term $[\varphi^I]_I * [\psi^J]_J$ in the above sum. It has three properties:

(a) $[\varphi^I]_I * [\psi^J]_J = [\theta^{I,J}]_K$ with $[\theta^{I,J}]_K$ a "bump" scaled according to K, where $K = I \vee J$; that is $K = (k_1, k_2, \ldots, k_n)$, and $k_m = \max(i_m, j_m)$, $1 \le m \le n$. This conclusion holds even if we do not assume the cancellation conditions on $[\varphi^I]_I$ and $[\psi^J]_J$.

(b) Next, because we do have the cancellation conditions (1.4) and (1.4′), we have a gain: There exists $\epsilon > 0$ so that $[\theta^{I,J}]_K$ can be written as a finite sum of terms of the form

$$\prod_{l \in A} 2^{-\epsilon|i_l - j_l|} \prod_{m \in B} 2^{-\epsilon[(i_{m+1} - i_m) + (j_{m+1} - j_m)]} \left[\widetilde{\theta}^{I,J}\right]_K$$

where $[\widetilde{\theta}^{I,J}]_K$ is another bump function scaled according to K, and A and B are disjoint sets with $A \cup B = \{1, \ldots, n\}$, with $n \notin B$.

(c) Strong cancellation fails in general for $[\theta^{I,J}]_K$, but weak cancellation holds.

Statements (a), (b), and (c) above are contained in Lemmas 6.17 and 6.18. With these assertions proved, one can proceed roughly[2] as follows. We define

$$\tilde{\theta}_K = \sum_{I \vee J = K} [\varphi^I]_I * [\varphi^J]_J,$$

where the sum is taken over all pairs (I, J) for which $I \vee J = K$. Because of the exponential gain given in (b) this sum converges to a K-scaled bump function. Moreover, because of (c), $\tilde{\theta}_K$ satisfies the weak cancellation property. As a result, the sum $\sum_K \tilde{\psi}_K$ converges to a flag kernel, and hence $\mathcal{K}_1 * \mathcal{K}_2$ is a flag kernel as was to be shown.

We comment briefly on the arguments needed to establish (b) and (c). Here we use the strong cancellation properties of $[\varphi^I]_I$ (or $[\psi^J]_J$). For (b) we express $[\varphi^I]_I$ as a sum of derivatives with respect to appropriate coordinates, then re-express these in terms of left-invariant vector fields, and finally pass these differentiation to $[\psi^J]_J$. The reverse may be done starting with cancellation of $[\psi^J]_J$. To obtain (c), the weak cancellation of $[\varphi^I]_I * [\psi^J]_J$, we begin the same way, but express $[\varphi^I]_I$ in terms of right-invariant vector fields and then pass these differentiations onto the resulting convolution products. The mechanism underlying this technique is set out in the various lemmas of Section 3.

The argument is a little more complex when we are in the case of equality for some of the indices that arise in I or J. This in effect involves convolutions with kernels belonging to coarser flags. The guiding principle for convolutions of such bump functions (or kernels) is that if $\mathcal{K}_j$ are flag kernels corresponding to the flags $\mathcal{F}_j$, $j = 1, 2$, then $\mathcal{K}_1 * \mathcal{K}_2$ is a flag kernel for the coarsest flag $\mathcal{F}$ that is finer than $\mathcal{F}_1$ and $\mathcal{F}_2$. The combinatorics involved are illustrated by several examples given in Sections 7.3 and 7.5.

---

[2] There are actually additional complications. We must first make a preliminary partition of the set of all pairs (I, J), and the result is that $\mathcal{K}_1 * \mathcal{K}_2$ is actually a finite sum of flag kernels.

1.5. L^p estimates via square functions

The proof of the L^p estimates (Theorem 8.14) starts with the descending chain of subgroups $G = G_1 \supset G_2 \supset \cdots \supset G_n$, where

$$G_m = \big\{ \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) : \mathbf{x}_1 = 0, \, \mathbf{x}_2 = 0, \ldots, \mathbf{x}_{m-1} = 0 \big\}$$

when $m \geq 2$. We observe that the dilations δ_r restrict to automorphisms of the G_m. We then proceed as follows:

(i) The standard (one-parameter) maximal functions and square functions on each group G_m, as given in [3], are then "lifted" (or "transferred") to the group G.

(ii) Compositions of these lifted objects lead to (n-parameter) maximal functions and square functions on G. Among these is the "strong" maximal function

$$M(f)(\mathbf{x}) = \sup \frac{1}{m(R_s)} \int_{R_s} |f(\mathbf{x}\mathbf{y}^{-1})| \, d\mathbf{y} \,,$$

for which one can prove vector-valued L^p inequalities. Here $R_s = \{\mathbf{x} : |\mathbf{x}_k| \leq s_k^{d_k}\}$, with $(s_1, \ldots, s_n)$ restricted to $s_1 \leq s_2 \leq \cdots \leq s_n$. There are also a pair of square functions, S and $\mathfrak{S}$, with the properties:

(1.8) $\| f \|_{L^p} \leq A_p \, \| S(f) \|_{L^p}$ and $\| \mathfrak{S}(f) \|_{L^p} \leq A'_p \, \| f \|_{L^p}$, for $1 < p < \infty$.

(iii) The connection of these square functions with our operators T, given by $Tf = f * \mathcal{K}$ with $\mathcal{K}$ a flag kernel, comes about because of the pointwise estimate:

(1.9) $$S(Tf)(\mathbf{x}) \leq c \, \mathfrak{S}(f)(\mathbf{x}) \,,$$

which is Lemma 8.13 below.

Now, (1.8), together with (1.9), prove the L^p boundedness of our operators.

Among the ideas used to prove (1.9) is the notion of a "truncated" flag kernel: such a kernel is truncated at "width a", $a \geq 0$, if it satisfies the conditions such as (1.1), but with $N_1 + \cdots + N_k$ replaced by $a + N_1 + \cdots + N_k$ throughout (see Definition 6.20). A key fact that is exploited is that a convolution of a bump of width b with a truncated kernel of width a yields a truncated kernel of width $a+b$. For this, see Theorem 8.7, and its consequence, Theorem 8.9.

1.6. Final remarks

The collection of operators with flag kernels contains both the automorphic (non-isotropic) Calderón–Zygmund operators as well as the usual isotropic Calderón–Zygmund operators with kernels of compact support (broadly speaking, the standard pseudo-differential operators of order 0). But flag kernels, by their definition, may have singularities away from the origin. Thus the algebra we are considering

consists of operators that are not necessarily pseudo-local. The study of a narrower algebra that arises naturally, which consists of pseudo-local operators and yet contains both types of Calderón–Zygmund operators, will be the subject of the second paper [10] in this series.

The authors are grateful to Brian Street for conversations and suggestions about the decomposition of flag kernels into sums of dilates of compactly supported functions. We would also like to thank the referee for a very careful reading of the paper. We note that the topic of this paper was the subject of several lectures given by one of us (EMS), in particular at a conference in honor of F. Treves at Rutgers, April, 2005, and at Washington University and UCLA in April and October 2008. During the preparation of this paper we learned of the work of Głowacki ([4], [5], and [6]) where overlapping results are obtained by different methods. We should also mention a forthcoming paper of Brian Street [12] that deals with the L^2-theory in a more general context than is done in the present paper.

2. Dilations and flag kernels on $\mathbb{R}^N$

Throughout the paper we shall use standard multi-index notation. $\mathbb{Z}$ denotes the set of integers, and $\mathbb{N}$, the set of non-negative integers. If $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^N$, then $|\alpha| = \alpha_1 + \cdots + \alpha_N$ and $\alpha! = \alpha_1! \cdots \alpha_N!$. If $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$, then $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_N^{\alpha_N}$. For $1 \leq j \leq N$, ∂_{x_j} (or more simply ∂_j) denotes the differential operator $\frac{\partial}{\partial x_j}$. If $\alpha \in \mathbb{N}^N$, then ∂^α denotes the partial differential operator $\partial_1^{\alpha_1} \cdots \partial_N^{\alpha_N}$.

The space of infinitely differentiable real-valued functions on $\mathbb{R}^N$ with compact support is denoted by $\mathcal{C}_0^\infty(\mathbb{R}^N)$ and the space of Schwartz functions is denoted by $\mathcal{S}(\mathbb{R}^N)$. The basic semi-norms on these spaces are defined as follows:

if $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, $\quad ||\varphi||_{(m)} = \sup \left\{ |\partial_{\mathbf{x}}^\alpha \varphi(\mathbf{x})| : |\alpha| \leq m, \ \mathbf{x} \in \mathbb{R}^N \right\}$;

if $\varphi \in \mathcal{S}(\mathbb{R}^N)$, $\quad ||\varphi||_{[M]} = \sup \left\{ |(1 + |\mathbf{x}|_e)^\alpha \partial_{\mathbf{x}}^\beta \varphi(\mathbf{x})| : |\alpha| + |\beta| \leq M, \mathbf{x} \in \mathbb{R}^N \right\}$.

Here $|\mathbf{x}|_e$ denotes the usual Euclidean length of $\mathbf{x} \in \mathbb{R}^N$.

2.1. The basic family of dilations

Fix positive real numbers $0 < d_1 \leq d_2 \leq \cdots \leq d_N$, and define a one-parameter family of dilations on $\mathbb{R}^N$ by setting

$$(2.1) \qquad \delta_r[\mathbf{x}] = r \cdot \mathbf{x} = \left(r^{d_1} x_1, \ldots, r^{d_N} x_N \right).$$

Also fix a smooth homogeneous norm $|\mathbf{x}|$ on $\mathbb{R}^N$ so that $|r \cdot \mathbf{x}| = r \, |\mathbf{x}|$. The homogeneous ball of radius r is $B(r) = \{ \mathbf{x} \in \mathbb{R}^N : |\mathbf{x}| < r \}$, and the homogeneous dimension of $\mathbb{R}^N$ (relative to this family of dilations) is $Q = d_1 + \cdots + d_N$. Recall that $|\mathbf{x}|_e = \sqrt{x_1^2 + \cdots + x_N^2}$ denotes the ordinary Euclidean length of a vector $\mathbf{x} \in \mathbb{R}^N$. If $m(\mathbf{x}) = c \mathbf{x}^\alpha = c x_1^{\alpha_1} \cdots x_N^{\alpha_N}$ is a monomial, then $m(r \cdot \mathbf{x}) = r^{\alpha_1 d_1 + \cdots + \alpha_N d_N} m(\mathbf{x})$,

and the homogeneous degree of m is $\Delta(m) = \alpha_1 d_1 + \cdots + \alpha_N d_N$. In particular, the homogeneous degree of a constant is zero. We shall agree that if the homogeneous degree of a monomial is negative, the monomial itself must be identically zero. With this convention, if m is any monomial, we have

$$(2.2) \qquad \Delta(\partial_j m) = \Delta(m) - d_j.$$

We denote by $\mathcal{H}_d$ the space of real-valued polynomials which are sums of monomials of homogeneous degree d. We have the following easy result:

Proposition 2.1. *If P is a polynomial, then $P \in \mathcal{H}_d$ if and only if $P(\mathbf{x}) = \sum_{\alpha \in \mathfrak{H}_d} c_\alpha \mathbf{x}^\alpha$, where $\mathfrak{H}_d = \big\{\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^n \mid \sum_{j=1}^N \alpha_j d_j = d\big\}$. Moreover,*

(1) *if $P \in \mathcal{H}_d$, then $P(r \cdot \mathbf{x}) = r^d P(\mathbf{x})$;*

(2) *if $P \in \mathcal{H}_{d_1}$ and $Q \in \mathcal{H}_{d_2}$, then $PQ \in \mathcal{H}_{d_1+d_2}$;*

(3) *if $P \in \mathcal{H}_d$, then $\partial_k(P)(\mathbf{x}) \equiv 0$ if $d_k > d$.*

2.2. Standard flags and flag kernels in $\mathbb{R}^N$

If X is an N-dimensional vector space, an *n-step* flag in X is a collection of subspaces $X_j \subseteq X$, $1 \le j \le n$, such that $(0) \subsetneq X_1 \subsetneq X_2 \subsetneq \cdots \subsetneq X_{n-1} \subsetneq X_n = X$. When $X = \mathbb{R}^N$ we single out a special class of *standard* flags. These are parameterized by partitions $N = a_1 + \cdots + a_n$ (where each a_j is a positive integer) as follows. We write

$$(2.3) \qquad \mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n},$$

and we write $\mathbf{x} \in \mathbb{R}^N$ as $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with $\mathbf{x}_j \in \mathbb{R}^{a_j}$. With an abuse of notation, we identify $\mathbb{R}^{a_k}$ with vectors in $\mathbb{R}^N$ of the form $(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{x}_k, \mathbf{0}, \ldots, \mathbf{0})$. Then the *standard flag* $\mathcal{F}$ associated to the partition $N = a_1 + \cdots + a_n$ and to the decomposition (2.3) is given by

$$(2.4) \quad (0) \subset \mathbb{R}^{a_n} \subset \mathbb{R}^{a_{n-1}} \oplus \mathbb{R}^{a_n} \subset \cdots \subset \mathbb{R}^{a_2} \oplus \cdots \oplus \mathbb{R}^{a_n} \subset \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n} = \mathbb{R}^N.$$

In dealing with such decompositions and flags, it is important to make clear which variables in $\mathbb{R}^N$ appear in which factor $\mathbb{R}^{a_l}$. We can write $\mathbf{x} \in \mathbb{R}^N$ either as $\mathbf{x} = (x_1, \ldots, x_N)$ with each $x_j \in \mathbb{R}$, or as $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ with $\mathbf{x}_l = (x_{p_l}, \ldots, x_{q_l}) \in \mathbb{R}^{a_l}$ so that $q_l = p_l + a_l - 1$. Denote by $J_l = \{p_l, p_l + 1, \ldots, q_l\}$ the set of subscripts corresponding to the factor $\mathbb{R}^{a_l}$ so that $\{1, \ldots, N\}$ is the disjoint union $J_1 \cup \cdots \cup J_n$. There is a mapping $\pi : \{1, \ldots, N\} \mapsto \{1, \ldots, n\}$ so that $j \in J_{\pi(j)}$ for $1 \le j \le N$. Thus for example $\pi(10) = 3$ means that the variable x_{10} belongs to the factor $\mathbb{R}^{a_3}$.

With the family of dilations defined in (2.1), the action on the subspace $\mathbb{R}^{a_l}$ is given by

$$(2.5) \qquad r \cdot \mathbf{x}_l = \big(r^{d_{p_l}} x_{p_l}, \ldots, r^{d_{q_l}} x_{q_l}\big).$$

The homogeneous dimension of $\mathbb{R}^{a_l}$ is

$$(2.6) \qquad Q_l = d_{p_l} + \cdots + d_{q_l} = \sum_{j \in J_l} d_j.$$

The function

$$(2.7) \qquad N_l(\mathbf{x}_l) = \sup_{p_l \leq s \leq q_l} |x_s|^{1/d_s}$$

is a homogeneous norm on $\mathbb{R}^{a_l}$ so that $N_l(r \cdot \mathbf{x}_l) = r \, N_l(\mathbf{x}_l)$. If $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^N$, let $\bar{\alpha}_l = (\alpha_{p_l}, \ldots, \alpha_{q_l})$, and set

$$(2.8) \qquad [\![\bar{\alpha}_l]\!] = \alpha_{p_l} d_{p_l} + \cdots + \alpha_{q_l} d_{q_l} = \sum_{j \in J_l} \alpha_j d_j.$$

We introduce a partial order on the set of all standard flags on $\mathbb{R}^N$.

Definition 2.2. Let $\mathcal{A} = (a_1, \ldots, a_r)$ and $\mathcal{B} = (b_1, \ldots, b_s)$ be two partitions of N so that $N = a_1 + \cdots + a_r = b_1 + \cdots + b_s$.

(1) The partition $\mathcal{A}$ is *finer* than the partition $\mathcal{B}$ (or $\mathcal{B}$ is *coarser* than $\mathcal{A}$) if there are integers $1 = \alpha_1 < \alpha_2 < \cdots < \alpha_{s+1} = r + 1$ so that $b_k = \sum_{j=\alpha_k}^{\alpha_{k+1}-1} a_j$. We write $\mathcal{A} \preceq \mathcal{B}$ or $\mathcal{B} \succeq \mathcal{A}$. If $\mathcal{A} \preceq \mathcal{B}$ but $\mathcal{A} \neq \mathcal{B}$ we write $\mathcal{A} \prec \mathcal{B}$ or $\mathcal{B} \succ \mathcal{A}$.

(2) If $\mathcal{F}_{\mathcal{A}}$ and $\mathcal{F}_{\mathcal{B}}$ are the flags corresponding to the two partitions and if $\mathcal{A} \preceq \mathcal{B}$ (or $\mathcal{A} \prec \mathcal{B}$), we say that the flag $\mathcal{F}_{\mathcal{A}}$ is finer than $\mathcal{F}_{\mathcal{B}}$ (or $\mathcal{F}_{\mathcal{B}}$ is coarser than $\mathcal{F}_{\mathcal{A}}$) and we also write $\mathcal{F}_{\mathcal{A}} \preceq \mathcal{F}_{\mathcal{B}}$ and $\mathcal{F}_{\mathcal{B}} \succeq \mathcal{F}_{\mathcal{A}}$ (or $\mathcal{F}_{\mathcal{A}} \prec \mathcal{F}_{\mathcal{B}}$ and $\mathcal{F}_{\mathcal{B}} \succ \mathcal{F}_{\mathcal{A}}$).

We recall from [9] the concept of a flag kernel on the vector space $\mathbb{R}^N$ associated to the decomposition $\mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$, equipped with the family of dilations given in equation (2.1). Let $\mathcal{F}$ be the standard flag given in (2.4). In order to formulate the cancellation conditions on the flag kernel, we need notation which allows us to split the variables $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ into two disjoint sets. Thus if $L = \{l_1, \ldots, l_\alpha\}$ and $M = \{m_1, \ldots, m_\beta\}$ are complementary subsets of $\{1, \ldots, n\}$ so that $\alpha + \beta = n$, let $N_a = a_{l_1} + \cdots + a_{l_\alpha}$ and $N_b = a_{m_1} + \cdots + a_{m_\beta}$. Write $\mathbf{x} \in \mathbb{R}^N$ as $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$ where $\mathbf{x}' = (\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_\alpha})$ and $\mathbf{x}'' = (\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta})$. If f is a function on $\mathbb{R}^{N_a}$ and g is a function on $\mathbb{R}^{N_b}$, define a function $f \otimes g$ on $\mathbb{R}^N$ by setting

$$f \otimes g(\mathbf{x}_1, \ldots, \mathbf{x}_n) = f(\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_\alpha}) \, g(\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}).$$

Definition 2.3. A *flag kernel* adapted to the flag $\mathcal{F}$ is a distribution $\mathcal{K} \in \mathcal{S}'(\mathbb{R}^N)$ which satisfies the following differential inequalities (part (a)) and cancellation conditions (part (b)):

(a) For test functions supported away from the subspace $\mathbf{x}_1 = 0$, the distribution $\mathcal{K}$ is given by integration against a $\mathcal{C}^\infty$-function K. Moreover, for every $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{Z}^N$ there is a constant C_α so that if $\bar{\alpha}_k = (\alpha_{p_k}, \ldots, \alpha_{q_k})$, then for $\mathbf{x}_1 \neq 0$,

$$\left| \partial^\alpha K(\mathbf{x}) \right| \leq C_\alpha \prod_{k=1}^n \left[N_1(\mathbf{x}_1) + \cdots + N_k(\mathbf{x}_k) \right]^{-Q_k - [\![\bar{\alpha}_k]\!]}.$$

(b) Let $\{1, \ldots, n\} = L \cup M$, with $L = \{l_1, \ldots, l_\alpha\}$, $M = \{m_1, \ldots, m_\beta\}$, and $L \cap M = \emptyset$ be any pair of complementary subsets. For any $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^{N_b})$ and any positive real numbers $R_1, \ldots, R_\beta$, put

$$\psi_R(\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}) = \psi(R_1 \cdot \mathbf{x}_{m_1}, \ldots, R_\beta \cdot \mathbf{x}_{m_\beta}).$$

Define a distribution $\mathcal{K}_{\psi,R}^\# \in \mathcal{S}'(\mathbb{R}^{a_{l_1} + \cdots + a_{l_r}})$ by setting

$$\langle \mathcal{K}_{\psi,R}^\#, \varphi \rangle = \langle \mathcal{K}, \psi_R \otimes \varphi \rangle$$

for any test function $\varphi \in \mathcal{S}(\mathbb{R}^{a_{l_1} + \cdots + a_{l_r}})$. Then the distribution $\mathcal{K}_{\psi,R}^\#$ satisfies the differential inequalities of part (a) for the decomposition $\mathbb{R}^{a_{l_1}} \oplus \cdots \oplus \mathbb{R}^{a_{l_r}}$. Moreover, the corresponding constants that appear in these differential inequalities are *independent* of the parameters $\{R_1, \ldots, R_s\}$, and depend only on the constants $\{C_\alpha\}$ from part (a), the semi-norms of ψ, and the support of ψ.

The constants $\{C_\alpha\}$ in part (a) and the implicit constants in part (b) are called the *flag kernel constants* for the flag kernel $\mathcal{K}$.

Remarks 2.4. (a) This definition proceeds by induction on the length n of the flag. The case $n = 1$ corresponds to Calderón–Zygmund kernels, and the inductive definition is invoked in part (b).

(b) With an abuse of notation, the distribution $\mathcal{K}_{\psi,R}^\#$ is often written

$$K_{\psi,R}^\#(\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_r}) = \int \cdots \int_{\mathbb{R}^{a_{m_1}} \oplus \cdots \oplus \mathbb{R}^{a_{m_s}}} K(\mathbf{x})\, \psi_R(\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_s})\, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_s}.$$

3. Homogeneous vector fields

In Section 6.6 below, where we consider a nilpotent Lie group G whose underlying space is $\mathbb{R}^N$, we will need to consider the families of left and right invariant vector fields on G. At this stage, before we introduce the group structure, we consider instead two spanning sets of vector fields $\{X_1, \ldots, X_N\}$ and $\{Y_1, \ldots Y_N\}$ on $\mathbb{R}^N$ which are homogeneous with respect to the basic family of dilations given in (2.1); this means that if Z_j is either X_j or Y_j for $1 \le j \le N$, then Z_j can be written[3]

$$(3.1) \qquad Z_j[\psi](\mathbf{x}) = \partial_j[\psi](\mathbf{x}) + \sum_{d_l > d_j} P_j^l(\mathbf{x}) \partial_l[\psi](\mathbf{x}),$$

with $P_j^l \in \mathcal{H}_{d_l - d_j}$. It follows from part (3) of Proposition 2.1 that $\partial_k(P_j^l) \equiv 0$ if $d_k > d_l$. Thus we can commute the operators given by multiplication by the polynomial P_j^l and differentiation with respect to x_l and also write

$$(3.2) \qquad Z_j[\psi](\mathbf{x}) = \partial_j[\psi](\mathbf{x}) + \sum_{d_l > d_j}^{N} \partial_l[P_j^l \psi](\mathbf{x}).$$

It follows from (3.1) or (3.2) that $Z_N = \partial_N$.

---

[3]Despite some risk of confusion, we do not introduce different notations for the coefficients of X_j and Y_j.

Proposition 3.1. *If $P \in \mathcal{H}_d$ and Z_j is either X_j or Y_j, then $Z_j[P] \in \mathcal{H}_{d-d_j}$, and if $d_j > d$, $Z_j[P] \equiv 0$.*

Proof. It follows from (2.2) that if $P \in \mathcal{H}_d$, then $\partial_l[P] \in \mathcal{H}_{d-d_l}$, and since $P_j^l \in \mathcal{H}_{d_l-d_j}$, it follows from part (2) of Proposition 2.1 that $P_j^l \, \partial_{x_l}[P] \in \mathcal{H}_{d-d_j}$. Thus $Z_j[P] \in \mathcal{H}_{d-d_j}$. The last conclusion then follows from part (3) of Proposition 2.1. $\square$

In equations (3.1) or (3.2), the vector fields $\{Z_j\}$ are written in terms of the Euclidean derivatives. Because these equations are in upper triangular form, it is easy to solve for the Euclidean derivatives in terms of the vector fields.

Proposition 3.2. *For each $1 \le j \le N$, let Z_j denote either X_j or Y_j. Then there are polynomials $Q_k^l \in \mathcal{H}_{d_l-d_k}$ such that, for $\psi \in \mathcal{S}(\mathbb{R}^N)$,*

$$\partial_k[\psi](\mathbf{x}) = Z_k[\psi](\mathbf{x}) + \sum_{d_l > d_k}^{N} Q_k^l(\mathbf{x}) \, Z_l[\psi](\mathbf{x}) = Z_k[\psi](\mathbf{x}) + \sum_{d_l > d_k}^{N} Z_l\big[Q_k^l \psi\big](\mathbf{x}).$$

Proof. We argue by reverse induction on the index k. When $k = N$ it follows from equation (3.1) that $\partial_N = Z_N = X_N = Y_N$. To establish the induction step, suppose that the conclusion of the proposition is true for all indices greater than k. From equation (3.1) and the induction hypothesis, for either choice of Z_k we have

$$\partial_k[\psi] = Z_k[\psi] - \sum_{d_m > d_k}^{N} P_k^m \partial_m[\psi] = Z_k[\psi] - \sum_{d_m > d_k}^{N} P_k^m \Big[Z_m[\psi] + \sum_{d_l > d_m}^{N} Q_m^l Z_l[\psi] \Big]$$

$$= Z_k[\psi] - \sum_{d_m > d_k}^{N} P_k^m Z_m[\psi] - \sum_{d_l > d_k} \Big[\sum_{d_k < d_m < d_l} P_k^m Q_m^l \Big] Z_l[\psi].$$

But according to part (2) of Proposition 2.1, $P_k^m \, Q_m^l \in \mathcal{H}_{d_l-d_k}$, and this completes the proof. $\square$

For $\psi \in \mathcal{S}(\mathbb{R}^N)$ and $t > 0$ set $\psi_t(\mathbf{x}) = \psi(t^{-1} \cdot \mathbf{x})$. Then multiplication by a polynomial $P \in \mathcal{H}_d$ is an operator homogeneous of degree d in the sense that $P(\mathbf{x})\psi_t(\mathbf{x}) = t^d (P\psi)_t(\mathbf{x})$, and the vector fields X_j and Y_j are operators homogeneous of degree $-d_j$ in the sense that $X_j[\varphi_t](\mathbf{x}) = t^{-d_j}(X_j\varphi)_t(\mathbf{x})$, $Y_j[\varphi_t](\mathbf{x}) = t^{-d_j}(Y_j\varphi)_t(\mathbf{x})$. In particular, the commutators $[X_j, X_k]$ and $[Y_j, Y_k]$ are vector fields which are homogeneous of degree $-(d_j + d_k)$. It follows that we can write

$$[X_j, X_k] = \sum_{d_m \ge d_j + d_k} Q_{j,k}^m(\mathbf{x}) \, \partial_m = \sum_{d_m \ge d_j + d_k} R_{j,k}^m(\mathbf{x}) \, Z_m, \text{ with } Q_{j,k}^m, R_{j,k}^m \in \mathcal{H}_{d_m-d_j-d_k},$$

$$[Y_j, Y_k] = \sum_{d_m \ge d_j + d_k} \widetilde{Q}_{j,k}^m(\mathbf{x}) \, \partial_m = \sum_{d_m \ge d_j + d_k} \widetilde{R}_{j,k}^m(\mathbf{x}) \, Z_m, \text{ with } \widetilde{Q}_{j,k}^m, \widetilde{R}_{j,k}^m \in \mathcal{H}_{d_m-d_j-d_k}.$$

If the operators $\{X_j\}$ and $\{Y_j\}$ are bases for a Lie algebra (as in the case of left or right invariant vector fields), the coefficients $\{R_{j,k}^m\}$ and $\{\widetilde{R}_{j,k}^m\}$ are constants.

Equations (3.1) or (3.2) express the vector fields $\{Z_k\}$ in terms of the standard derivatives $\{\partial_j\}$, and Proposition 3.2 expresses the standard derivatives in terms of the vector fields. We shall need analogous identities for products of r vector fields $Z_{k_1} \cdots Z_{k_r}$ or products of r Euclidean derivatives $\partial_{k_1} \cdots \partial_{k_r}$. The formulas are somewhat complicated, since they involve products of operators of various lengths. To help with the formulation of the results, it will be convenient to introduce the following notation:

Definition 3.3. *Let* $k_1, \ldots, k_r \in \{1, \ldots, N\}$ *be a set of* r *integers, possibly with repetitions.*

(1) *For any non-empty set* $U \subset \{1, \ldots, r\}$, *put* $d_U = \sum_{\ell \in U} d_{k_\ell}$ *and*

$$\mathfrak{I}(U) = \{m \in \{1, \ldots, N\} : d_m \geq d_U\}.$$

Note that if U *consists of two or more elements and* $m \in \mathfrak{I}(U)$, *then* $d_m > \sup_{\ell \in U} d_{k_\ell}$.

(2) *For each integer* $1 \leq s \leq r$, *let* $\mathcal{U}_s^r$ *denote the set of partitions of the set* $\{1, \ldots, r\}$ *into* s *non-empty disjoint subsets* $\mathcal{U} = \{U_1, \ldots, U_s\}$.

The following proposition then shows how to write products of vector fields in terms of products of Euclidean derivatives.

Proposition 3.4. *Let* $k_1, \ldots, k_r \in \{1, \ldots, N\}$ *be a set of* r *integers, possibly with repetitions. For* $1 \leq \ell \leq r$, *let* Z_{k_ℓ} *denote either* X_{k_ℓ} *or* Y_{k_ℓ}. *Then there are polynomials* $P_{U_\ell}^{m_\ell} \in \mathcal{H}_{d_{m_\ell} - d_{U_\ell}}$ *such that*

$$Z_{k_1} \cdots Z_{k_r}[\psi] = \sum_{s=1}^{r} \sum_{(U_1, \ldots, U_s) \in \mathcal{U}_s^r} \sum_{m_1 \in \mathfrak{I}(U_1)} \cdots \sum_{m_s \in \mathfrak{I}(U_s)} \partial_{m_1} \cdots \partial_{m_s} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \psi \right].$$

If $s = r$, *so that* $U_j = \{k_j\}$, *the polynomial* $P_{U_j}^{k_j}(\mathbf{x}) \equiv 1$.

Proof. We argue by induction on r. The case $r = 1$ is contained in equation (3.2), so suppose we are given vector fields $\{Z_{k_1}, \ldots, Z_{k_{r+1}}\}$ where each Z_{k_ℓ} is either X_{k_ℓ} or Y_{k_ℓ}. Then, by induction,

$$Z_{k_1} \cdots Z_{k_{r+1}}[\psi] = Z_{k_1} \cdots Z_{k_r} \left[Z_{k_{r+1}}[\psi] \right]$$

$$(3.3) \quad = \sum_{s=1}^{r} \sum_{(U_1, \ldots, U_s) \in \mathcal{U}_s^r} \sum_{m_1 \in \mathfrak{I}(U_1)} \cdots \sum_{m_s \in \mathfrak{I}(U_s)} \partial_{m_1} \cdots \partial_{m_s} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \left[Z_{k_{r+1}}[\psi] \right] \right].$$

Since we can write

$$Z_{k_{r+1}}[\psi] = \partial_{k_{r+1}}[\psi] + \sum_{\{m \,:\, d_m > d_{k_{r+1}}\}} \partial_m \left[P_{k_{r+1}}^m \psi \right]$$

where $P_{k_{r+1}}^m \in \mathcal{H}_{d_m - d_{k_{r+1}}}$, the derivative $\partial_{m_1} \cdots \partial_{m_s} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} [Z_{k_{r+1}}[\psi]] \right]$ in the

last line of (3.3) can be written

$$
\begin{aligned}
\partial_{m_1} \cdots \partial_{m_s} & \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \left[Z_{k_{r+1}}[\psi] \right] \right] \\
&= \partial_{m_1} \cdots \partial_{m_s} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \left[\partial_{k_{r+1}}[\psi] \right] \right] \\
&\quad + \sum_{d_m > d_{k_{r+1}}} \partial_{m_1} \cdots \partial_{m_s} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \left[\partial_m [P_{k_{r+1}}^m \psi] \right] \right] \\
&= \partial_{m_1} \cdots \partial_{m_s} \partial_{k_{r+1}} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \psi \right] \\
&\quad - \partial_{m_1} \cdots \partial_{m_s} \left[\partial_{k_{r+1}} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \right] \psi \right] \\
&\quad + \sum_{d_m > d_{k_{r+1}}} \partial_{m_1} \cdots \partial_{m_s} \partial_m \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} P_{k_{r+1}}^m \psi \right] \\
&\quad - \sum_{d_m > d_{k_{r+1}}} \partial_{m_1} \cdots \partial_{m_s} \left[\partial_m \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \right] P_{k_{r+1}}^m \psi \right].
\end{aligned}
$$

The terms in the first and third lines of the last expression have the right form for the case $r + 1$. Thus in the term $\partial_{m_1} \cdots \partial_{m_s} \partial_{k_{r+1}} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \psi \right]$, r has been replaced by $r + 1$, and the set $\{1, \ldots, r, r+1\}$ has been decomposed into $s + 1$ subsets $\{U_1, \ldots, U_s, U_{s+1}\}$ where $U_{s+1} = \{k_{r+1}\}$, and $P_{U_{s+1}}^{k_{r+1}}(\mathbf{x}) \equiv 1$. The same is true for each term $\partial_{m_1} \cdots \partial_{m_s} \partial_m \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} P_{k_{r+1}}^m \psi \right]$, except that $P_{U_{s+1}}^m = P_{k_{r+1}}^m$.

For the terms in the second and fourth lines, we use the product rule; we write $\partial_{k_{r+1}} \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \right]$ and $\partial_m \left[P_{U_1}^{m_1} \cdots P_{U_s}^{m_s} \right]$ as a sum of s terms. For example,

$$
\begin{aligned}
\sum_{d_m > d_{k_{r+1}}} \partial_{m_1} \cdots \partial_{m_s} & \left[\partial_m \left[P_{U_1}^{m_1} \right] P_{U_2}^{m_2} \cdots P_{U_s}^{m_s} P_{k_{r+1}}^m \psi \right] \\
&= \partial_{m_1} \cdots \partial_{m_s} \left[\left(\sum_{d_m > d_{k_{r+1}}} \partial_m \left[P_{U_1}^{m_1} \right] P_{k_{r+1}}^m \right) P_{U_2}^{m_2} \cdots P_{U_s}^{m_s} \psi \right] \\
&= \partial_{m_1} \cdots \partial_{m_s} \left[\widetilde{P}_{\widetilde{U}_1}^{m_1} P_{U_2}^{m_2} \cdots P_{U_s}^{m_s} \psi \right]
\end{aligned}
$$

where $\widetilde{P}_{\widetilde{U}_1}^{m_1} = \sum_{d_m > d_{k_{r+1}}} \partial_m \left[P_{U_1}^{m_1} \right] P_{k_{r+1}}^m \in \mathcal{H}_{d_{m_1} - d_{U_1} - d_{k_{r+1}}}$. These terms also have the right form for the case $r + 1$, since we now let $\widetilde{U}_1 = U_1 \cup \{k_{r+1}\}$, so that $\{k_1, \ldots, k_{r+1}\} = \widetilde{U}_1 \cup U_2 \cup \cdots \cup U_s$. This establishes the proposition.   □

The next result shows how to write products of Euclidean derivatives in terms of products of vector fields. Since the proof is similar to that of Proposition 3.4, we omit it.

Proposition 3.5. *Let $k_1, \ldots, k_r \in \{1, \ldots, N\}$ be a set of r integers, possibly with repetitions, chosen from the set $\{1, \ldots, N\}$. There are polynomials $Q_{U_\ell}^{m_\ell} \in \mathcal{H}_{d_{m_\ell} - d_{U_\ell}}$ such that*

$$
\partial_{k_1} \cdots \partial_{k_r}[\psi] = \sum_{s=1}^{r} \sum_{(U_1, \ldots, U_s) \in \mathcal{U}_s^r} \sum_{m_1 \in \mathfrak{I}(U_1)} \cdots \sum_{m_s \in \mathfrak{I}(U_s)} Z_{m_1} \cdots Z_{m_s} [Q_{U_1}^{m_1} \cdots Q_{U_s}^{m_s} \psi].
$$

Here each Z_j is either X_j or Y_j. If $s = r$, so that $U_\ell = \{k_\ell\}$, the polynomial $Q_{U_\ell}^{m_\ell}(\mathbf{x}) \equiv 1$.

4. Normalized bump functions and their dilations

4.1. Families of dilations

Fix the family of dilations on $\mathbb{R}^N$ given in equation (2.1). We introduce an N-*para-meter* family of dyadic[4] dilations. For $f \in L^1(\mathbb{R}^N)$ and $I \in \mathbb{Z}^N$ set

(4.1)
$$2^I \cdot \mathbf{x} = (2^{-d_1 i_1} x_1, \ldots, 2^{-d_N i_N} x_N),$$

$$[f]_I(\mathbf{x}) = 2^{-\sum\limits_{\ell=1}^{N} d_\ell i_\ell} f(2^I \cdot \mathbf{x}).$$

Then $\big|\big|[f]_I\big|\big|_{L^1(\mathbb{R}^N)} = ||f||_{L^1(\mathbb{R}^N)}$. The set of monotone increasing indices is denoted by

(4.2)
$$E_N = \left\{ I = (i_1, \ldots, i_N) \in \mathbb{Z}^N \,\big|\, i_1 \leq i_2 \leq \cdots \leq i_N \right\}.$$

When we consider flag kernels corresponding to the decomposition $\mathfrak{A}$ given by $N = a_1 + \cdots + a_n$ and $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$, we consider the n-parameter family of dilations parameterized by n-tuples $I = (i_1, \ldots, i_n) \in \mathbb{Z}^n$:

(4.3)
$$2^I \cdot \mathbf{x} = (2^{i_1} \cdot \mathbf{x}_1, \ldots, 2^{i_n} \cdot \mathbf{x}_n), \quad \text{where}$$

$$2^{i_\ell} \cdot \mathbf{x}_\ell = (2^{-d_{p_\ell} i_\ell} x_{p_\ell}, \ldots, 2^{-d_{q_\ell} i_\ell} x_{q_\ell}), \quad \text{and}$$

$$[f]_I(\mathbf{x}) = 2^{-\sum\limits_{\ell=1}^{n} Q_\ell i_\ell} f(2^I \cdot \mathbf{x}).$$

The set of monotone increasing indices in this case is denoted by

(4.4)
$$\mathcal{E}_n = \left\{ I = (i_1, \ldots, i_n) \in \mathbb{Z}^n \,\big|\, i_1 \leq i_2 \leq \cdots \leq i_n \right\}.$$

Given the decomposition $\mathfrak{A}$, there is a mapping $p_{\mathfrak{A}} : \mathcal{E}_n \mapsto E_N$ given by

(4.5)
$$p_{\mathfrak{A}}(i_1, \ldots, i_n) = (\overbrace{i_1, \ldots, i_1}^{a_1}, \overbrace{i_2, \ldots, i_2}^{a_2}, \ldots, \overbrace{i_n, \ldots, i_n}^{a_n}).$$

We shall want to write flag kernels as sums of dilates $[\varphi]_I$ of normalized bump functions φ. Roughly speaking, a family of functions $\{\varphi_\alpha\}$ in $\mathcal{C}_0^\infty(\mathbb{R}^N)$ or $\mathcal{S}(\mathbb{R}^N)$ is *normalized* if one has uniform control of the supports (in the case of $\mathcal{C}_0^\infty(\mathbb{R}^N)$) and of the semi-norms $||\varphi_\alpha||_{(m)}$ or $||\varphi_\alpha||_{[M]}$. The following definition will simplify the precise statements of our results.

Definition 4.1. (1) If $\varphi, \widetilde{\varphi} \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, then $\widetilde{\varphi}$ is *normalized in terms of* φ if[5] there are constants $C, C_m > 0$ and integers $p_m \geq 0$ so that:

(a) If the support of φ is contained in the ball $B(\rho)$, then the support of $\widetilde{\varphi}$ is contained in the ball $B(C\rho)$.

(b) For every non-negative integer m, $||\widetilde{\varphi}||_{(m)} \leq C_m ||\varphi||_{(m+p_m)}$.

---

[4]In Section 8, we shall use a continuous version of this dyadic family. If $\mathbf{t} = (t_1, \ldots, t_N)$ with each $t_j > 0$, we will set $f_{\mathbf{t}}(\mathbf{x}) = f(t_1^{d_1} x_1, \ldots, t_N^{d_N} x_N)$.

[5]We shall sometimes use the expressions 'normalized with respect to' or 'normalized relative to' as a variant of 'normalized in terms of'.

(2) If $\psi, \widetilde{\psi} \in \mathcal{S}(\mathbb{R}^n)$, then $\widetilde{\psi}$ is *normalized in terms of* ψ if there are constants $C_N > 0$ and integers $p_N \geq 0$ so that $||\widetilde{\psi}||_{[N]} \leq C_N \, ||\psi||_{[N+p_N]}$ for every non-negative integer N.

(3) If $P, \widetilde{P}$ are polynomials, then $\widetilde{P}$ is normalized in terms of P if $\widetilde{P}$ is obtained from P by multiplying each coefficient by a constant of modulus less than or equal to 1.

If $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, it is sometimes convenient to write φ as a sum of products of functions of a single variable. That this is possible follows from the following fact:

Proposition 4.2. *Let* $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$. *Then for each* $\alpha \in \mathbb{N}^N$ *and* $1 \leq k \leq N$ *there are functions* $\varphi_{\alpha,k} \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ *so that*

$$\varphi(x_1, \ldots, x_N) = \sum_{\alpha \in \mathbb{N}^N} c_\alpha \varphi_{\alpha,1}(x_1) \cdots \varphi_{\alpha,n}(x_n),$$

where for any $M > 0$, *there is a constant* C_M *such that* $|c_\alpha| \leq C_M \, (1 + |\alpha|)^{-M}$.

Proposition 4.2 follows easily by appropriately periodizing φ, expanding φ in a rapidly converging Fourier series, and then multiplying by appropriate cutoff functions $\psi(x_1) \cdots \psi(x_N)$.

4.2. Differentiation and multiplication of dilates of bump functions

In this section we study the action of differentiation or multiplication by a homogeneous polynomial on dilates of bump functions. The key results are Proposition 4.7 and Corollary 4.8 below. We begin with the following result, which follows easily from the definitions and the chain rule:

Proposition 4.3. *Let* $\psi \in \mathcal{S}(\mathbb{R}^N)$. *Then*

$$(4.6) \qquad 2^{+d_k i_k} \, \partial_k [\psi]_I(\mathbf{x}) = [\partial_k \psi]_I(\mathbf{x}) \quad and \quad 2^{-d_k i_k} \, x_k \, [\psi]_I(\mathbf{x}) = [x_k \, \psi]_I(\mathbf{x}).$$

More generally, if $I = (i_1, \ldots, i_N) \in \mathbb{Z}^N$, *and* $\alpha \in \mathbb{N}^N$, *then*

$$(4.7) \qquad 2^{+[\![\alpha \cdot I]\!]} \partial^\alpha [\psi]_I = \left[\partial^\alpha \psi \right]_I \quad and \quad 2^{-[\![\alpha \cdot I]\!]} \mathbf{x}^\alpha [\psi]_I = \left[\mathbf{x}^\alpha \psi \right]_I,$$

where $[\![\alpha \cdot I]\!] = \sum_{k=1}^N \alpha_k d_k i_k$.

We will frequently use the following generalization of the second identity in equation (4.7):

Proposition 4.4. *Let* $P \in \mathcal{H}_d$ *where* $d < d_l$. *If* $I \in E_N$, *there is a polynomial* $P_I \in \mathcal{H}_d$, *normalized in terms of* P, *so that*

$$P(\mathbf{x})[\psi]_I(\mathbf{x}) = 2^{d \, i_l} \left[P_I \psi \right]_I (\mathbf{x})$$

for $\psi \in \mathcal{S}(\mathbb{R}^N)$.

Proof. Write $P(\mathbf{x}) = \sum_{\alpha \in \mathfrak{H}_d} c_\alpha \mathbf{x}^\alpha$ with

$$\mathfrak{H}_d = \{\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^N : \alpha_1 d_1 + \cdots + \alpha_N d_N = d\}.$$

Since $d < d_l$, if $\alpha \in \mathfrak{H}_d$ we have $\alpha_j = 0$ for $j \geq l$. According to Proposition 4.3, it follows that $P(\mathbf{x})[\psi]_I(\mathbf{x}) = \sum_{\alpha \in \mathfrak{H}_d} c_\alpha 2^{[\![\alpha \cdot I]\!]} [\mathbf{x}^\alpha \psi]_I(\mathbf{x})$, and

$$[\![\alpha \cdot I]\!] = \sum_{m=1}^{l-1} \alpha_m d_m i_m = i_l \sum_{m=1}^{l-1} \alpha_m d_m - \sum_{m=1}^{l-1} \alpha_m d_m(i_l - i_m) = d\, i_l - \sum_{m=1}^{l-1} \alpha_m d_m(i_l - i_m).$$

Thus

$$P(\mathbf{x})\big[\psi\big]_I(\mathbf{x}) = 2^{d\, i_l} \Big[\sum_{\alpha \in \mathfrak{H}_d} c_\alpha 2^{-\sum_{m=1}^{l-1} \alpha_m d_m(i_l - i_m)} \mathbf{x}^\alpha \psi \Big]_I(\mathbf{x}).$$

Since $I \in E_N$, each exponent $-\sum_{m=1}^{l-1} \alpha_m d_m(i_l - i_m) \leq 0$. The proof is complete if we set $P_I(\mathbf{x}) = \sum_{\alpha \in \mathfrak{H}_d} c_\alpha 2^{-\sum_{m=1}^{l-1} \alpha_m d_m(i_l - i_m)} \mathbf{x}^\alpha$. $\square$

Remark. Equation (4.6) shows that the operator $2^{d_k i_k} \partial_{x_k}$, applied to the I-dilate of a function φ, is the I-dilate of a function $\widetilde{\varphi}$ normalized in terms of φ, and multiplying the I-dilate of function φ by $2^{-d_k i_k} x_k$ is the I-dilate of a function $\widetilde{\varphi}$ normalized in terms of φ. Thus at 'scale I', the operators $2^{d_k i_k} \partial_{x_k}$ and $2^{-d_k i_k} x_k$ are 'invariant'; they map the collection of I-dilates of normalized functions to itself.

A key observation, which is used when we consider convolution on homogeneous nilpotent groups, is that we can replace the operator $2^{d_k i_k} \partial_{x_k}$ with the operator $2^{d_k i_k} Z_k$, or conversely the operator $2^{d_k i_k} Z_k$ by the operator $2^{d_k i_k} \partial_{x_k}$, at the cost of introducing an error involving terms $2^{d_l i_l} Z_l$ or $2^{d_l i_l} \partial_l$, where $l > k$, multiplied by a 'gain' $2^{-d_k(i_l - i_k)}$. The precise statement is given in Proposition 4.5 below.

Let $P_k^l \in \mathcal{H}_{d_l - d_k}$ be the homogeneous polynomials that are coefficients of a vector field Z_k as in equations (3.1) or (3.2), and let $Q_k^l \in \mathcal{H}_{d_l - d_k}$ be the polynomials in Proposition 3.2. Since $d_l - d_k < d_l$, we can use Proposition 4.4 to write

$$(4.8) \qquad \begin{aligned} P_k^l(\mathbf{x})[\psi]_I(\mathbf{x}) &= 2^{i_l(d_l - d_k)} \big[P_{k,I}^l \psi\big]_I(\mathbf{x}), \\ Q_k^l(\mathbf{x})[\psi]_I(\mathbf{x}) &= 2^{i_l(d_l - d_k)} \big[Q_{k,I}^l \psi\big]_I(\mathbf{x}), \end{aligned}$$

where $P_{k,I}^l$, $Q_{k,I}^l \in \mathcal{H}_{d_l - d_k}$ are normalized relative to P_k^l and Q_k^l

Proposition 4.5. *Let $I \in E_N$, and let $\{P_{k,I}^l\}$ and $\{Q_{k,I}^l\}$ be the homogeneous polynomials defined in equation (4.8). Let $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ (respectively $\psi \in \mathcal{S}(\mathbb{R}^N)$). Then,*

$$(2^{d_k i_k} Z_k)[\psi]_I = (2^{d_k i_k} \partial_k)[\psi]_I + \sum_{d_l > d_k}^{N} 2^{-d_k(i_l - i_k)} (2^{d_l i_l} \partial_l)\big[P_{k,I}^l \psi\big]_I,$$

$$(2^{d_k i_k} \partial_k)[\psi]_I = (2^{d_k i_k} Z_k)[\psi]_I + \sum_{d_l > d_k}^{N} 2^{-d_k(i_l - i_k)} (2^{d_l i_l} Z_l)\big[Q_{k,I}^l \psi\big]_I.$$

The functions $\{P^l_{k,I}\psi\}$ *and* $\{Q^l_{k,I}\psi\}$ *are normalized with respect to* ψ *in* $\mathcal{C}^\infty_0(\mathbb{R}^N)$ *(respectively in* $\mathcal{S}(\mathbb{R}^N)$*).*

Proof. The first identity follows immediately from equations (3.2) and (4.8). To obtain the second, use Proposition 3.2 and equation (4.8); we have

$$2^{d_k i_k}\partial_{x_k}[\psi]_I = 2^{d_k i_k}Z_k[\psi]_I + 2^{d_k i_k}\sum_{l=k+1}^{N}Z_lQ^l_k(\mathbf{x})[\psi]_I$$

$$= 2^{d_k i_k}Z_k[\psi]_I + \sum_{l=k+1}^{N}2^{-d_k(i_l-i_k)}(2^{d_l i_l}Z_l)\big[Q^l_{k,I}\psi\big]_I,$$

which is the desired formula. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 4.6. *If* $\varphi \in \mathcal{C}^\infty_0(\mathbb{R}^N)$ *(respectively* $\varphi \in \mathcal{S}(\mathbb{R}^N)$*) and if* $I \in E_N$*, then there is a function* $\widetilde{\varphi} \in \mathcal{C}^\infty_0(\mathbb{R}^N)$ *(respectively* $\widetilde{\varphi} \in \mathcal{S}(\mathbb{R}^N)$*), normalized with respect to* φ*, such that* $2^{d_k i_k}Z_k[\varphi]_I = [\widetilde{\varphi}]_I$*.*

We shall need an analogue of Proposition 4.5 for r-fold products of vector fields or Euclidean derivatives. If $\{P^{m_l}_{U_l}\}$ and $\{Q^{m_l}_{U_l}\}$ are the polynomials appearing in Propositions 3.4 and 3.5, we use Proposition 4.4 to define polynomials $P^{m_l}_{U_l,I}$ and $Q^{m_l}_{U_l,I}$ by the formulas

(4.9)
$$P^{m_l}_{U_l}(\mathbf{x})[\psi]_I(\mathbf{x}) = 2^{i_l(d_{m_l}-d_{U_l})}\big[P^{m_l}_{U_l,I}\psi\big]_I(\mathbf{x}),$$
$$Q^{m_l}_{U_l}(\mathbf{x})[\psi]_I(\mathbf{x}) = 2^{i_l(d_{m_l}-d_{U_l})}\big[Q^{m_l}_{U_l,I}\psi\big]_I(\mathbf{x}).$$

Proposition 4.7. *Let* $k_1,\ldots,k_r \in \{1,\ldots,N\}$ *be a set of* r *integers, possibly with repetitions, and let* $I \in E_N$*. Let* Z_ℓ *denote either* X_ℓ *or* Y_ℓ*. Then*

$$(2^{d_{k_1}i_{k_1}}Z_{k_1})\cdots(2^{d_{k_r}i_{k_r}}Z_{k_r})[\psi]_I$$

$$= \sum_{s=1}^{r}\sum_{(U_1,\ldots,U_s)\in\mathcal{U}^r_s}\sum_{m_1\in\mathfrak{I}(U_1)}\cdots\sum_{m_s\in\mathfrak{I}(U_s)}2^{\sum_{\ell=1}^{r}d_{k_\ell}i_{k_\ell}-\sum_{\ell=1}^{s}i_{m_\ell}d_{U_\ell}}$$

$$\cdot(2^{d_{m_1}i_{m_1}}\partial_{m_1})\cdots(2^{d_{m_s}i_{m_s}}\partial_{m_s})\big[P^{m_1}_{U_1,I}\cdots P^{m_s}_{U_s,I}\psi\big]_I,$$

and

$$(2^{d_{k_1}i_{k_1}}\partial_{k_1})\cdots(2^{d_{k_r}i_{k_r}}\partial_{k_r})[\psi]_I$$

$$= \sum_{s=1}^{r}\sum_{(U_1,\ldots,U_s)\in\mathcal{U}^r_s}\sum_{m_1\in\mathfrak{I}(U_1)}\cdots\sum_{m_s\in\mathfrak{I}(U_s)}2^{\sum_{\ell=1}^{r}d_{k_\ell}i_{k_\ell}-\sum_{\ell=1}^{s}i_{m_\ell}d_{U_\ell}}$$

$$\cdot(2^{d_{m_1}i_{m_1}}Z_{m_1})\cdots(2^{d_{m_s}i_{m_s}}Z_{m_s})\big[Q^{m_1}_{U_1,I}\cdots Q^{m_s}_{U_s,I}\psi\big]_I.$$

In either identity, if $s = r$ *so that* $U_\ell = \{k_\ell\}$*,* $1 \le \ell \le r$*, the polynomials*

$$P^{k_\ell}_{U_\ell}(\mathbf{x}) = Q^{k_\ell}_{U_\ell}(\mathbf{x}) \equiv 1.$$

Proof. Using Proposition 3.4, we have

$$
(2^{d_{k_1} i_{k_1}} \partial_{k_1}) \cdots (2^{d_{k_r} i_{k_r}} \partial_{k_r})[\psi]_I
$$

$$
= 2^{\sum_{\ell=1}^{r} d_{k_\ell} i_{k_\ell}} \sum_{s=1}^{r} \sum_{\mathcal{U} \in \mathcal{U}_s^r} \sum_{j=1}^{s} \sum_{m_j \in \mathfrak{I}(U_j)} Z_{m_1} \cdots Z_{m_s} \left[Q_{U_1}^{m_1} \cdots Q_{U_s}^{m_s} [\psi]_I \right]
$$

$$
= 2^{\sum_{\ell=1}^{r} d_{k_\ell} i_{k_\ell}} \sum_{s=1}^{r} \sum_{\mathcal{U} \in \mathcal{U}_s^r} \sum_{j=1}^{s} \sum_{m_j \in \mathfrak{I}(U_j)} 2^{\sum_{\ell=1}^{s} i_{m_\ell}(d_{m_\ell} - d_{U_\ell})} Z_{m_1} \cdots Z_{m_s} \left[Q_{U_1,I}^{m_1} \cdots Q_{U_s,I}^{m_s} \psi \right]_I
$$

$$
= \sum_{s=1}^{r} \sum_{\mathcal{U} \in \mathcal{U}_s^r} \sum_{j=1}^{s} \sum_{m_j \in \mathfrak{I}(U_j)} 2^{\sum_{\ell=1}^{r} d_{k_\ell} i_{k_\ell} - \sum_{\ell=1}^{s} i_{m_\ell} d_{U_\ell}}
$$

$$
\cdot (2^{d_{m_1} i_{m_1}} Z_{m_1}) \cdots (2^{d_{m_s} i_{m_s}} Z_{m_s}) \left[Q_{U_1,I}^{m_1} \cdots Q_{U_s,I}^{m_s} \psi \right]_I .
$$

This completes the proof of the first identity. The second identity is established in the same way. □

In Proposition 4.7, we can rewrite the exponent of the power of 2 as follows. Having chosen $\mathcal{U} = (U_1, \ldots, U_s) \in \mathcal{U}_s^r$, there is a unique mapping $\sigma = \sigma_{\mathcal{U}} : \{1, \ldots, r\} \to \{1, \ldots, s\}$ so that $\ell \in U_{\sigma(\ell)}$ for $1 \le \ell \le r$. Then

$$
\sum_{\ell=1}^{r} d_{k_\ell} i_{k_\ell} - \sum_{\ell=1}^{s} i_{m_\ell} d_{U_\ell} = \sum_{\ell=1}^{r} d_{k_\ell} i_{k_\ell} - \sum_{\ell=1}^{r} i_{m_{\sigma(\ell)}} d_{k_\ell} = -\sum_{\ell=1}^{r} d_{k_\ell}(i_{m_{\sigma(\ell)}} - i_{k_\ell}).
$$

Since $\ell \in U_{\sigma(\ell)}$ and $m_{\sigma(\ell)} \in \mathfrak{I}(U_{\sigma(\ell)})$, it follows that $d_{m_{\sigma(\ell)}} \ge d_{U_{\sigma(\ell)}} \ge d_{k_\ell}$, with equality only possible if $U_{\sigma(\ell)} = \{\sigma(\ell)\}$. Thus if $I \in E_N$, it follows that $i_{m_{\sigma(\ell)}} \ge i_{k_\ell}$, in which case

$$
2^{\sum_{\ell=1}^{r} d_{k_\ell} i_{k_\ell} - \sum_{\ell=1}^{s} i_{m_\ell} d_{U_\ell}} \le 2^{-\epsilon \sum_{\ell=1}^{r} (i_{m_{\sigma(\ell)}} - i_{k_\ell})},
$$

where we can take $\epsilon = d_1 > 0$.

We can now recast the identities in Proposition 4.7 in a way which, although losing some information, makes them more useful and easier to work with when dealing with flag kernels.

Corollary 4.8. *Fix a decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$. Let $k_1, \ldots, k_r \in \{1, \ldots, N\}$ be a set of r integers, possibly with repetitions, with $k_\ell \in J_{\pi(\ell)}$.[6] For $1 \le \ell \le r$, let Z_{k_ℓ} denote either X_{k_ℓ} or Y_{k_ℓ}. Let $I \in \mathcal{E}_n$, and let $\psi \in \mathcal{S}(\mathbb{R}^N)$ or $\mathcal{C}_0^\infty(\mathbb{R}^N)$.*

(1) *The function $(2^{d_{k_1} i_{\pi(1)}} Z_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_r})[\psi]_I$ can be written as a finite sum of terms of the following form. Decompose the set $\{1, \ldots, r\}$ into two disjoint complementary subsets A and B with $\pi(j) \ne n$ for any $j \in A$ and $B \ne \emptyset$. Let*

---

[6]Recall from page 638 that $\pi : \{1, \ldots, N\} \to \{1, \ldots, n\}$; for any coordinate x_j, $1 \le j \le N$, then x_j is a coordinate in the factor $\mathbb{R}^{a_{\pi(j)}}$, and $J_{\pi(j)}$ is the set of indices of all the coordinates in $\mathbb{R}^{a_{\pi(j)}}$.

$B = \{\ell_1, \dots, \ell_s\}$, *and choose integers* $\bar{m} = \{m_1, \dots, m_s\}$ *so that each* $m_t \in J_{\ell_t}$. *Then a typical term in the expansion of* $(2^{d_{k_1} i_{\pi(1)}} Z_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_r})[\psi]_I$ *is*

$$2^{-\epsilon \sum_{j \in A}(i_{\pi(j)+1} - i_{\pi(j)})} (2^{d_{m_1} i_{\pi(1)}} \partial_{m_1}) \cdots (2^{d_{m_s} i_{\pi(s)}} \partial_{m_s}) [\psi_{A,B,\bar{m}}]_I,$$

where $\psi_{A,B,\bar{m}}$ *is normalized relative to* ψ.

(2) *The function* $(2^{d_{k_1} i_{\pi(1)}} \partial_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} \partial_{k_r})[\psi]_I$ *can be written as a finite sum of terms of the following form. Decompose the set* $\{1, \dots, r\}$ *into two disjoint complementary subsets* A *and* B *with* $\pi(j) \neq n$ *for any* $j \in A$ *and* $B \neq \emptyset$. *Let* $B = \{\ell_1, \dots, \ell_s\}$, *and choose integers* $\bar{m} = \{m_1, \dots, m_s\}$ *so that each* $m_t \in J_{\ell_t}$. *Then a typical term in the expansion of* $(2^{d_{k_1} i_{\pi(1)}} \partial_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} \partial_{k_r})[\psi]_I$ *is*

$$2^{-\epsilon \sum_{j \in A}(i_{\pi(j)+1} - i_{\pi(j)})} (2^{d_{m_1} i_{\pi(1)}} Z_{m_1}) \cdots (2^{d_{m_s} i_{\pi(s)}} Z_{m_s}) [\psi_{A,B,\bar{m}}]_I,$$

where $\psi_{A,B,\bar{m}}$ *is normalized relative to* ψ.

Remarks 4.9. (a) The essential point of part (1) in the corollary is that, when replacing the operator

$$(2^{d_{k_1} i_{\pi(1)}} Z_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_r})[\psi]_I$$

with a sum of terms of the form

$$2^{-\epsilon \sum_{j \in A}(i_{\pi(j)+1} - i_{\pi(j)})} (2^{d_{m_1} i_{\pi(1)}} \partial_{m_1}) \cdots (2^{d_{m_s} i_{\pi(s)}} \partial_{m_s}) [\psi_{A,B,\bar{m}}]_I,$$

either the factor $(2^{d_{k_\ell} i_{\pi(\ell)}} Z_{k_\ell})$ is replaced by a term $(2^{d_{m_\ell} i_{\pi(\ell)}} \partial_{m_\ell})$, where the coordinate x_{m_ℓ} belongs to the *same* subspace as x_{k_ℓ} and hence has the same dilation $i_{\pi(\ell)}$, or it is replaced by the gain $2^{-\epsilon(i_{\pi(\ell)+1} - i_{\pi(\ell)})}$. Part (2) is the same assertion with the roles of the vector fields and Euclidean differentiation interchanged.

(b) One term in the expansion of $(2^{d_{k_1} i_{\pi(1)}} Z_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_r})[\psi]_I$ in part (1) arises by letting $A = \emptyset$ so that $B = \{1, \dots, r\}$, and then choosing $m_r = k_r$. We have seen that in this case the function $\psi_{A,B,\bar{m}} = \psi$, so we get the term

$$(2^{d_{k_1} i_{\pi(1)}} \partial_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} \partial_{k_r})[\psi]_I.$$

Every other term in the expansion then involves either a gain $2^{-\epsilon(i_{\pi(\ell)+1} - i_{\pi(\ell)})}$ or the replacement of a variable x_{k_ℓ} by a different variable x_{m_ℓ} with $d_{m_\ell} > d_{k_\ell}$. We shall say that any term of the form

$$2^{-\epsilon \sum_{j \in A}(i_{\pi(j)+1} - i_{\pi(j)})} (2^{d_{m_1} i_{\pi(1)}} \partial_{m_1}) \cdots (2^{d_{m_s} i_{\pi(s)}} \partial_{m_s}) [\psi_{A,B,\bar{m}}]_I,$$

where either $A \neq \emptyset$ or some $d_{m_\ell} > d_{k_\ell}$ is an *allowable error*. Thus the difference

$$(2^{d_{k_1} i_{\pi(1)}} Z_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_r})[\psi]_I - (2^{d_{k_1} i_{\pi(1)}} \partial_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} \partial_{k_r})[\psi]_I$$

is a sum of allowable errors.

(c) Since Euclidean derivatives commute, it follows that if σ is any permutation of the set $\{1, \dots, r\}$, then the difference

$$(2^{d_{k_1} i_{\pi(1)}} Z_{k_1}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_r})[\psi]_I - (2^{d_{k_1} i_{\pi(1)}} Z_{k_{\sigma(1)}}) \cdots (2^{d_{k_r} i_{\pi(r)}} Z_{k_{\sigma(r)}})[\psi]_I$$

is a sum of allowable errors.

5. Cancellation

A function is often said to have *cancellation* if its average or integral is zero. For our purposes, we shall need a more refined notion involving integrals in some subset of variables. Let $J = \{j_1, \ldots, j_s\} \subseteq \{1, \ldots, N\}$. If $\psi \in \mathcal{S}(\mathbb{R}^N)$, write

$$(5.1) \qquad \int \psi(\mathbf{x}) \, d\mathbf{x}_J = \int_{\mathbb{R}^s} \psi(x_1, \ldots, x_N) \, dx_{j_1} \cdots dx_{j_s}.$$

Note that $\int \psi(\mathbf{x}) \, d\mathbf{x}_J$ is then a function of the variables x_k for which $k \notin J$. We say that a function ψ has *cancellation in the variables* $\{x_{j_1}, \ldots, x_{j_s}\}$ if $\int_{\mathbb{R}^s} \psi(\mathbf{x}) \, d\mathbf{x}_J \equiv 0$, where $J = \{j_1, \ldots, j_s\}$. Later, in Section 5.2, we shall give additional definitions of 'strong cancellation' and 'weak cancellation' for a function ψ.

5.1. Cancellation and the existence of primitives

We begin by showing that cancellation in certain collections of variables is equivalent to the existence of appropriate primitives.

Lemma 5.1. *Let* $\psi \in \mathcal{S}(\mathbb{R}^N)$, *and let* $J_k \subseteq \{1, \ldots, n\}$ *be non-empty subsets for* $1 \leq k \leq r$. *If the sets* $\{J_k\}$ *are mutually disjoint, then the following two statements are equivalent:*

(a) *For* $1 \leq k \leq r$, $\int \psi(\mathbf{x}) \, d\mathbf{x}_{J_k} = 0$.

(b) *There are functions* $\psi_{j_1,\ldots,j_r} \in \mathcal{S}(\mathbb{R}^N)$, *normalized with respect to* ψ, *such that*

$$\psi(\mathbf{x}) = \sum_{j_1 \in J_1} \cdots \sum_{j_r \in J_r} \partial_{j_1} \cdots \partial_{j_r} \psi_{j_1,\ldots,j_r}(\mathbf{x}).$$

Moreover, if the function $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, *then we can choose the functions* $\psi_{j_1,\ldots,j_r} \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ *normalized with respect to* ψ.

It follows easily from the Fundamental Theorem of Calculus that (b) implies (a). The main content of the lemma is thus the opposite implication. This will follow by induction on r from the following assertion:

Proposition 5.2. *Let* $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$. *Suppose that* $J_1, J_2 \subset \{1, \ldots, N\}$ *are non-empty and disjoint. If* $\int \psi(\mathbf{x}) \, d\mathbf{x}_{J_1} = \int \psi(\mathbf{x}) \, d\mathbf{x}_{J_2} = 0$, *then for each* $k \in J_1$ *there is a function* $\psi_k \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, *normalized relative to* ψ, *so that:*

(i) *we can write* $\psi = \sum_{k \in J_1} \partial_k \psi_k$;

(ii) *for each* $k \in J_1$ *we still have* $\int \psi_k(\mathbf{x}) \, d\mathbf{x}_{J_2} = 0$.

If $\psi \in \mathcal{S}(\mathbb{R}^N)$, *the same conclusions hold except that the functions* $\psi_k \in \mathcal{S}(\mathbb{R}^N)$ *and are normalized relative to* ψ.

Proof. By relabeling the coordinates, we can assume that $J_1 = \{1, \ldots, k\}$, and that $J_2 \subseteq \{k + 1, \ldots, N\}$. Suppose that ψ has compact support in the set $B = \{\mathbf{x} \in \mathbb{R}^N : |x_j| < a_j\}$. Choose $\chi \in \mathcal{C}_0^\infty(\mathbb{R})$ with support in $[-1, +1]$ such that $\int_\mathbb{R} \chi(s) \, ds = 1$, and put $\chi_j(t) = a_j^{-1}\chi(a_j^{-1}t)$ so that χ_j is supported in $[-a_j, +a_j]$ and still has integral equal to 1. Put

$$\varphi_1(\mathbf{x}) = \psi(\mathbf{x}) - \chi_1(x_1) \int_\mathbb{R} \psi(s, x_2, \ldots, x_N) \, ds$$

and for $2 \le j \le k$, put

$$\varphi_j(\mathbf{x}) = \Big[\prod_{l=1}^{j-1} \chi_l(x_l) \Big] \int_{\mathbb{R}^{j-1}} \psi(s_1, \ldots, s_{j-1}, x_j, \ldots, x_N) \, ds_1 \cdots ds_{j-1}$$

$$- \Big[\prod_{l=1}^{j} \chi_l(x_l) \Big] \int_{\mathbb{R}^j} \psi(s_1, \ldots, s_j, x_{j+1}, \ldots, x_N) \, ds_1 \cdots ds_j.$$

Then the functions $\{\varphi_j\}$ have the following properties. First, since $\int \psi(\mathbf{x}) \, d\mathbf{x}_{J_1} = 0$, the second integral in the definition of the last function φ_k is zero, and hence $\psi(\mathbf{x}) = \sum_{j=1}^k \varphi_j(\mathbf{x})$. Next, it is clear that each φ_j is supported in the set B. Finally, for $1 \le j \le k$,

$$\int_\mathbb{R} \varphi_j(x_1, \ldots, x_{j-1}, s, x_{j+1}, \ldots, x_N) \, ds = 0,$$

so if we put

(5.2)
$$\psi_j(\mathbf{x}) = \int_{-\infty}^{x_j} \varphi_j(x_1, \ldots, x_{j-1}, s, x_{j+1}, \ldots, x_N) \, ds$$
$$= - \int_{x_j}^{\infty} \varphi_j(x_1, \ldots, x_{j-1}, s, x_{j+1}, \ldots, x_N) \, ds,$$

then ψ_j is supported on the set B, and $\varphi_j(\mathbf{x}) = \partial_j \psi_j(\mathbf{x})$. It is clear that one can estimate the size of the derivatives of the functions $\{\psi_j\}$ in terms of the derivatives of ψ, so ψ_j is normalized in terms of ψ. Moreover, since $\int \psi(\mathbf{x}) \, d\mathbf{x}_{J_2} = 0$, it follows from their definitions that $\int \varphi_j(\mathbf{x}) \, d\mathbf{x}_{J_2} = 0$, and hence $\int \psi_j(\mathbf{x}) \, d\mathbf{x}_{J_2} = 0$. This completes the proof if $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$. If $\psi \in \mathcal{S}(\mathbb{R}^N)$, the proof goes the same way. One only has to observe from equation (5.2) that the functions $\psi_j \in \mathcal{S}(\mathbb{R}^N)$. This completes the proof of Proposition 5.2, and hence Lemma 5.1 is also established. $\quad\square$

5.2. Strong and weak cancellation

In this section we introduce two kinds of cancellation conditions that can be imposed on functions in $\mathcal{S}(\mathbb{R}^N)$ or $\mathcal{C}_0^\infty(\mathbb{R}^N)$. As discussed in Section 1.1, these concepts are used in the context of the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$ given in (2.3), where $a_1 + \cdots + a_n = N$ and each $a_j \ge 1$. Recall that if $\mathbf{x} \in \mathbb{R}^N$, we write

$\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ where $\mathbf{x}_l = (x_{p_l}, x_{p_l+1}, \ldots, x_{q_l}) \in \mathbb{R}^{a_l}$. We then let J_l denote the set of integers $\{p_l, p_l + 1, \ldots, q_l\}$. If $\varphi \in \mathcal{S}(\mathbb{R}^N)$, set

$$(5.3) \qquad \int_{\mathbb{R}^{a_l}} \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \, d\mathbf{x}_l = \int_{\mathbb{R}^{a_l}} \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \, dx_{p_l} \cdots dx_{q_l}.$$

Definition 5.3. Fix the decomposition (2.3), and let $\varphi \in \mathcal{S}(\mathbb{R}^N)$. The function φ has *strong cancellation* if and only if

$$\int_{\mathbb{R}^{a_l}} \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \, d\mathbf{x}_\ell \equiv 0, \quad 1 \le \ell \le n.$$

That is, φ has strong cancellation if and only if it has cancellation in each collection of variables $\{x_{p_l}, \ldots, x_{q_l}\}$ for $1 \le l \le n$.

Remark 5.4. It follows from Proposition 5.2 that φ has strong cancellation if and only if there are functions $\varphi_{j_1,\ldots,j_n}$ normalized with respect to φ so that

$$\varphi = \sum_{j_1 \in J_1} \cdots \sum_{j_n \in J_n} \partial_{j_1} \cdots \partial_{j_n} \varphi_{j_1,\ldots,j_n}.$$

We now introduce a weaker cancellation condition.

Definition 5.5. Fix the decomposition (2.3) of $\mathbb{R}^N$. Let $\varphi \in \mathcal{S}(\mathbb{R}^N)$ or $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, and let $I = (i_1, \ldots i_n) \in \mathcal{E}_n$. The function φ has *weak cancellation* with parameter $\epsilon > 0$ relative to the multi-index I if and only if

$$\varphi = \sum_{\substack{A \cup B = \{1,\ldots,n\} \\ A = \{\alpha_1,\ldots,\alpha_r\} \\ n \notin B}} \left(\prod_{s \in B} 2^{-\epsilon(i_{s+1} - i_s)} \right) \sum_{j_1 \in J_{\alpha_1}} \cdots \sum_{j_r \in J_{\alpha_r}} \left(\partial_{j_1} \cdots \partial_{j_r} \right) [\varphi_{A,B,j_1,\ldots,j_r}]$$

where each $\varphi_{A,B,j_1,\ldots,j_r}$ is normalized relative to φ. Here the outer sum is taken over all decompositions of the set $\{1, \ldots, n\}$ into two disjoint subsets A and B such that $n \in A$.

According to Remark 5.4, φ has strong cancellation if it can be written as a sum of functions of the form $\partial_{j_1} \cdots \partial_{j_n} \varphi$; *i.e.*, as n^{th}-derivatives of functions where there is one derivative in a variable from each of the n subspaces $\mathbb{R}^{a_l}$. Definition 5.5 imposes a weaker condition; a function φ has weak cancellation if again it is a sum of terms, but the term $\partial_{j_1} \cdots \partial_{j_n} \varphi_{j_1,\ldots,j_n}$ is itself replaced by a new sum of terms. If a derivative ∂_{j_l} does not appear, so that the term does not have cancellation in the subspace $\mathbb{R}^{a_l}$, there is instead a gain given by $2^{-\epsilon(i_{l+1} - i_l)}$.

Remarks 5.6. There will be occasions when we will use the fact that a function φ has weak cancellation with respect to $I \in \mathcal{E}_n$ to draw inferences about existence of primitives and smallness in only some of the subspaces $\{\mathbb{R}^{a_l}\}$. In particular, the following assertions follow easily from Definition 5.5:

(1) Suppose that $I \in \mathcal{E}_n$ and that φ has weak cancellation relative to I. Let $M \subset \{1, \ldots, n\}$. Then

$$\varphi = \sum_{\substack{A \cup B = M \\ A = \{\alpha_1, \ldots, \alpha_r\} \\ n \notin B}} \left(\prod_{s \in B} 2^{-\epsilon(i_{s+1} - i_s)} \right) \sum_{j_1 \in J_{\alpha_1}} \cdots \sum_{j_r \in J_{\alpha_r}} (\partial_{j_1} \cdots \partial_{j_s}) [\widetilde{\varphi}_{A,B,j_1,\ldots,j_r}]$$

where each $\widetilde{\varphi}_{A,B,j_1,\ldots,j_r}$ is normalized relative to φ, and where the outer sum is over all subsets $B \subset M$, with the understanding that if $n \in M$ then $n \notin B$.

(2) In particular, if we take $M = \{1\}$, it follows that if φ has weak cancellation relative to $I \in \mathcal{E}_n$, we can write

$$\varphi = \sum_{r=1}^{a_1} \partial_r [\varphi_r] + 2^{-\epsilon(i_2 - i_1)} \varphi_0$$

where $\{\varphi_0, \varphi_1, \ldots, \varphi_{a_1}\}$ are normalized relative to φ. Here, the derivatives are with respect to the variables in the first subspace $\mathbb{R}^{a_1}$.

We can also characterize weak cancellation in terms of the "smallness" of integrals, to be compared with the characterization of strong cancellation in terms of the vanishing of integrals given in Lemma 5.1. For any partition $\{1, \ldots, n\} = A \cup B$ with $A = \{j_1, \ldots, j_a\}$ and $B = \{k_1, \ldots, k_b\}$, write $\mathbf{x} \in \mathbb{R}^N$ as $\mathbf{x} = (\mathbf{x}_A, \mathbf{x}_B)$ where $\mathbf{x}_A = (\mathbf{x}_{j_1}, \ldots, \mathbf{x}_{j_a})$, $\mathbf{x}_B = (\mathbf{x}_{k_1}, \ldots, \mathbf{x}_{k_b})$. Let $d\mathbf{x}_A = d\mathbf{x}_{j_1} \cdots d\mathbf{x}_{j_a}$, and let $d\mathbf{x}_B = d\mathbf{x}_{k_1} \cdots d\mathbf{x}_{k_b}$.

Proposition 5.7. *Let $\epsilon > 0$ and $I \in \mathcal{E}_n$. A function $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ has weak cancellation with parameter ϵ relative to I if and only if for every partition $\{1, \ldots, n\} = A \cup B$ into disjoint subsets we have*

$$\int_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \psi(\mathbf{x}_A, \mathbf{x}_B) \, d\mathbf{x}_B = \begin{cases} 0 & \text{if } n \in B, \\ \left[\prod_{k \in B} 2^{-\epsilon(i_{k+1} - i_k)} \right] \psi_A(\mathbf{x}_A) & \text{if } n \notin B, \end{cases}$$

where $\psi_A \in \mathcal{S}(\bigoplus_{j \in A} \mathbb{R}^{a_j})$ is normalized relative to ψ. If $\psi \in C_0^\infty(\mathbb{R}^N)$, the functions $\psi_A \in C_0^\infty(\bigoplus_{j \in A} \mathbb{R}^{a_j})$.

Proof. It is clear that if ψ has weak cancellation, then it satisfies the condition of Proposition 5.7, so the main content is the opposite implication.

Let $\chi_l \in \mathcal{C}_0^\infty(\mathbb{R}^{a_l})$ have support in the set where $|\mathbf{x}_l| \leq 1$, with $\int_{\mathbb{R}^{a_l}} \chi(\mathbf{x}_l) \, d\mathbf{x}_l = 1$. For $\psi \in \mathcal{S}(\mathbb{R}^N)$ and $1 \leq l \leq N$, define

$$L_l[\psi](\mathbf{x}) = \psi(\mathbf{x}) - \chi_l(\mathbf{x}_l) \int_{\mathbb{R}^{a_l}} \psi(\mathbf{x}) \, d\mathbf{x}_l,$$

$$M_l[\psi](\mathbf{x}) = \chi_l(\mathbf{x}_l) \int_{\mathbb{R}^{a_l}} \psi(\mathbf{x}) \, d\mathbf{x}_l.$$

It is easy to check that the operators $\{L_1, \ldots, L_n, M_1, \ldots, M_n\}$ all commute, and that

$$\int_{\mathbb{R}^{a_l}} M_l[\psi](\mathbf{x}) \, d\mathbf{x}_l = \int_{\mathbb{R}^{a_l}} \psi(\mathbf{x}) \, d\mathbf{x}_l \quad \text{and} \quad \int_{\mathbb{R}^{a_l}} L_l[\psi](\mathbf{x}) \, d\mathbf{x}_l = 0.$$

Since $L_l + M_l$ is the identity operator, if $\psi \in \mathcal{S}(\mathbb{R}^N)$,

$$\psi(\mathbf{x}) = \prod_{l=1}^{n} (L_l + M_l)[\psi](\mathbf{x})$$

$$= \sum_{A \cup B = \{1,\ldots,n\}} \Big(\prod_{j \in A} L_l \Big) \Big(\prod_{k \in B} M_k \Big)[\psi](\mathbf{x}) = \sum_{A \cup B = \{1,\ldots,n\}} \psi_{A,B}(\mathbf{x}).$$

It is clear that the functions $\{\psi_{A,B}\}$ are normalized relative to ψ.

Now suppose that ψ satisfies the hypotheses of Proposition 5.7. Since

$$\Big(\prod_{k \in B} M_k \Big)[\psi](\mathbf{x}) = \Big(\prod_{k \in B} \chi_k(\mathbf{x}_k) \Big) \int_{\oplus_{k \in B} \mathbb{R}^{a_k}} \psi(\mathbf{x}_A, \mathbf{x}_B) \, d\mathbf{x}_B,$$

it follows that $\psi_{A,B} \equiv 0$ if $n \in B$. Also, for every $j \in A$ we have

$$\int_{\mathbb{R}^{a_j}} \psi_{A,B}(\mathbf{x}) \, d\mathbf{x}_j = 0.$$

Thus for every decomposition $\{1,\ldots,n\} = A \cup B$ with $n \in A$, it follows that

$$\Big(\prod_{k \in B} M_k \Big)[\psi](\mathbf{x}) = \Big[\prod_{k \in B} 2^{-\epsilon(i_{k+1} - i_k)} \Big] \psi_A(\mathbf{x}_A) \Big(\prod_{k \in B} \chi_k(\mathbf{x}_k) \Big)$$

where ψ_A is normalized relative to ψ. But then

$$\psi_{A,B}(\mathbf{x}) = \Big(\prod_{j \in A} L_l \Big) \Big(\prod_{k \in B} M_k \Big)[\psi](\mathbf{x})$$

$$= \Big[\prod_{k \in B} 2^{-\epsilon(i_{k+1} - i_k)} \Big] \Big(\prod_{j \in A} L_l \Big)[\psi_A](\mathbf{x}_A) \Big(\prod_{k \in B} \chi_k(\mathbf{x}_k) \Big).$$

Now the function $\big(\prod_{j \in A} L_l \big)[\psi_A](\mathbf{x}_A) \big(\prod_{k \in B} \chi_k(\mathbf{x}_k) \big)$ has cancellation in each of the subspaces $\mathbb{R}^{a_j}$ for $j \in A$. If we write $A = \{\alpha_1, \ldots \alpha_r\}$, it follows from Lemma 5.1 that this function can be written as a sum of terms of the form $\partial_{j_1} \cdots \partial_{j_r}[\psi_{A,B,\sigma}]$, where the variable x_{j_k} belongs to the subspace $\mathbb{R}^{a_{\alpha_k}}$. This then makes it clear that ψ has weak cancellation, and completes the proof.   □

6. The structure of flag kernels

In this section we establish four important properties of flag kernels. The first (Theorem 6.1) shows how to decompose a given flag kernel into a sum of dilates of compactly supported functions with strong cancellation. The second (Theorem 6.8) shows conversely that a sum of dilates of Schwartz functions with weak cancellation converges to a flag kernel. The third (Theorem 6.13) shows that the cancellation conditions imposed in part (b) of Definition 2.3 can be relaxed. The fourth (Theorem 6.15) shows that the family of flag kernels is invariant under appropriately chosen changes of variables.

6.1. Dyadic decomposition of flag kernels

Every flag kernel is a product kernel in the sense of Definition 2.1.1 of [9]. (This is proved in [9]). Then Corollary 2.2.2 in that reference implies that if $\mathcal{K}$ is a flag kernel, then there are functions $\varphi^I \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, $I \in \mathbb{Z}^n$ with strong cancellation and uniformly bounded seminorms $||\varphi^I||_{(m)}$, supported in the set where $\frac{1}{2} \leq N_j(\mathbf{x}_j) \leq 4$ for every j, such that $\mathcal{K} = \sum_{I \in \mathbb{Z}^n} [\varphi^I]_I$ in the sense of distributions. Recall that $\mathcal{E}_n = \{I = (i_1, \ldots, i_n) \in \mathbb{Z}^n : i_1 \leq i_2 \leq \cdots \leq i_n\}$. Since a flag kernel satisfies better differential inequalities than a general product kernel, one expects that it should be possible to write $\mathcal{K}$ as a sum of dilates of functions $\varphi^I \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ where the dilations range *only* over the set $\mathcal{E}_n$ instead of over all of $\mathbb{Z}^n$. Such a result is stated in Corollary 2.2.4 of [9], but the precise statement there is not correct because one must allow additional terms involving flag kernels adapted to coarser flags. This section provides a correct statement and proof.

Theorem 6.1. *Let $\mathcal{K}$ be a flag kernel adapted to the standard flag $\mathcal{F}$*

$$(2.4) \quad (0) \subseteq \mathbb{R}^{a_n} \subseteq \mathbb{R}^{a_{n-1}} \oplus \mathbb{R}^{a_n} \subseteq \cdots \subseteq \mathbb{R}^{a_3} \oplus \cdots \oplus \mathbb{R}^{a_n} \subseteq \mathbb{R}^{a_2} \oplus \cdots \oplus \mathbb{R}^{a_n} \subseteq \mathbb{R}^N$$

Then there is a decomposition $\mathcal{K} = \mathcal{K}_0 + \mathcal{K}_1 + \cdots + \mathcal{K}_n$ with the following properties:

(1) *For each $I \in \mathcal{E}_n$ there is a function $\varphi^I \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ so that $\mathcal{K}_0 = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$ with convergence in the sense of distributions. Moreover,*

    (a) *the support of each function φ^I is contained in the unit ball $B = \{\mathbf{x} \in \mathbb{R}^N : |\mathbf{x}| \leq 1\}$;*

    (b) *there are constants $C_m > 0$ (depending on the constants for the flag kernel $\mathcal{K}$) so that for each $I \in \mathcal{E}_n$ and all $m \geq 0$, $||\varphi^I||_{(m)} \leq C_m$;*

    (c) *each function φ^I has strong cancellation in the sense of Definition 5.3: for $1 \leq j \leq n$,*

$$\int_{\mathbb{R}^{a_j}} \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_n) \, d\mathbf{x}_j = 0;$$

    (d) *for each $I \in \mathcal{E}_n$, $\varphi^I(\mathbf{x}_1, \ldots, \mathbf{x}_n) = 0$ for $|\mathbf{x}_1| < \frac{1}{8}$.*

(2) *For $1 \leq j \leq n$, each $\mathcal{K}_j$ is a flag kernel adapted to a flag which is* strictly coarser *than $\mathcal{F}$.*

Recall that in Section 1.2 we distinguished between elements $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$ where we have strict inequality $i_1 < i_2 < \cdots < i_{n-1} < i_n$, and elements where some of the entries are equal. Underlying this dichotomy is the fact that $\mathcal{E}_n$ is a polyhedral cone in $\mathbb{Z}^n$, the set of elements with strict inequality form the 'open' n-dimensional interior, and elements with various sets of equal elements correspond to lower dimensional faces. In Theorem 6.1, we may as well assume that the dyadic sum representing $\mathcal{K}_0$ extends over the interior n-tuples with strictly increasing components, leaving the "boundary n-tuples" with repeated indices to parametrize the dyadic terms of the sums representing the other $\mathcal{K}_j$'s.

Using induction on the number of steps in a flag, Theorem 6.1 immediately gives us the following corollary:

Corollary 6.2. *Let $\mathcal{K}$ be a flag kernel adapted to a standard flag $\mathcal{F}$ of length n. There is a finite collection of flags $\{\mathcal{F}_{k,s}\}$, $1 \leq k \leq n$ and $1 \leq s \leq b_k$, with the following properties:*

(1) *For $k = n$, $b_n = 1$ and $\mathcal{F}_{n,1} = \mathcal{F}$.*

(2) *For each $k < n$, the flag $\mathcal{F}_{k,s}$ has length k and is strictly coarser than $\mathcal{F}$.*

(3) *For each (k,s) there is a uniformly bounded family of functions $\{\varphi_{k,s}^J\} \subset \mathcal{C}_0^\infty(\mathbb{R}^N)$, $J \in \mathcal{E}_k$, all supported in the unit ball and having strong cancellation relative to the decomposition of $\mathbb{R}^N$ corresponding to the flag $\mathcal{F}_{k,s}$ so that in the sense of distributions,*

$$(6.1) \qquad \mathcal{K} = \sum_{k=1}^n \sum_{s=1}^{b_k} \sum_{J \in E_k} \left[\varphi_{k,s}^J\right]_J.$$

The proof of Theorem 6.1 relies on three preliminary results. The first, which gives the characterization of flag kernels in terms of their Fourier transforms, was established in [9]. We briefly recall the relevant definitions. Let $(\mathbb{R}^N)^*$ denote the space of linear functionals on $\mathbb{R}^N$, and for a subspace $W \subseteq \mathbb{R}^N$, let $W^\perp \subset (\mathbb{R}^N)^*$ be the subspace of linear functions which are zero on W. If $\mathcal{F}$ is the flag in $\mathbb{R}^N$ given in (2.4), the dual flag is $\mathcal{F}^*$ given by

$$(0)^\perp \supseteq (\mathbb{R}^{a_n})^\perp \supseteq (\mathbb{R}^{a_{n-1}} \oplus \mathbb{R}^{a_n})^\perp \supseteq \cdots \supseteq (\mathbb{R}^{a_2} \oplus \cdots \oplus \mathbb{R}^{a_n})^\perp = (\mathbb{R}^N)^\perp.$$

If we identify $(\mathbb{R}^{a_k} \oplus \cdots \oplus \mathbb{R}^{a_n})^\perp$ with $(\mathbb{R}^{a_1})^* \oplus \cdots \oplus (\mathbb{R}^{a_{k-1}})^*$, the dual flag becomes

$$(6.2) \quad (0) \subseteq (\mathbb{R}^{a_1})^* \subseteq (\mathbb{R}^{a_1})^* \oplus (\mathbb{R}^{a_2})^* \subseteq \cdots \subseteq (\mathbb{R}^{a_1})^* \oplus \cdots \oplus (\mathbb{R}^{a_{n-1}})^* \subseteq (\mathbb{R}^N)^*.$$

If $\xi \in (\mathbb{R}^N)^*$, we write $\xi = (\xi_1, \ldots, \xi_n)$ where $\xi_j = (\xi_{j,1}, \ldots, \xi_{j,a_j}) \in (\mathbb{R}^{a_j})^*$. The family of dilations on $\mathbb{R}^N$ defined in equation (2.1) induce a family of dilations on $(\mathbb{R}^N)^*$ so that $\langle r \cdot \mathbf{x}, \xi \rangle = \langle \mathbf{x}, r \cdot \xi \rangle$. We let $|\xi|$ be a smooth homogeneous norm on $(\mathbb{R}^N)^*$, and if $\xi_j \in (\mathbb{R}^{a_j})^*$, we let $|\xi_j|$ be the restriction of the norm to this subspace.

Definition 6.3. A *flag multiplier* relative to the flag $\mathcal{F}^*$ given in (6.2) is a function $m(\xi)$ which is infinitely differentiable away from the subspace $\xi_n = 0$, and which satisfies the differential inequalities

$$\left|\partial_{\xi_1}^{\bar{\alpha}_1} \cdots \partial_{\xi_n}^{\bar{\alpha}_n} m(\xi)\right| \leq C_\alpha \prod_{j=1}^n \left(|\xi_j| + \cdots + |\xi_n|\right)^{-[\![\bar{\alpha}_j]\!]}.$$

We can now state Theorem 2.3.9 of [9] as follows:

Lemma 6.4. *Let $\mathcal{K}$ be a flag kernel adapted to the flag $\mathcal{F}$. Then the Fourier transform of $\mathcal{K}$ is a flag multiplier relative to the dual flag $\mathcal{F}^*$. Conversely, every flag multiplier relative to the flag $\mathcal{F}^*$ is the Fourier transform of a flag kernel adapted to the flag $\mathcal{F}$.*

The next preliminary result provides a decomposition of test functions in $\mathcal{S}(\mathbb{R}^N)$.

Lemma 6.5. *Let $\psi \in \mathcal{S}(\mathbb{R}^N)$. Then there are functions $\{\psi^k\} \subset \mathcal{C}_0^\infty(\mathbb{R}^N)$, $k = 0, 1, 2, \ldots$, such that*

$$\psi(\mathbf{x}) = \sum_{k=0}^\infty 2^{-kQ} \psi^k(2^{-k} \cdot \mathbf{x}),$$

where Q is the homogeneous dimension of $\mathbb{R}^N$. Moreover these functions have the following properties:

(a) *each ψ^k is supported in the unit ball;*

(b) *for any $\delta > 0$ and any $\alpha \in \mathbb{N}^N$, there exists $M \in \mathbb{N}$ depending on δ and α so that*

$$\sup_{\mathbf{x} \in \mathbb{R}^N} \left| \partial^\alpha \psi^k(\mathbf{x}) \right| \leq ||\psi||_{[M]} \, 2^{-k\delta} \, ;$$

(c) *if ψ has strong cancellation, then each ψ^k has strong cancellation.*

Proof. Choose $\eta \in \mathcal{C}_0^\infty(\mathbb{R})$ supported in $[-1, -\frac{1}{8}] \cup [\frac{1}{8}, 1]$ with $\eta(t) = \eta(-t)$ such that

$$\sum_{k=-\infty}^{+\infty} \eta(2^{-k}t) = 1$$

for all $t \neq 0$. For any $t \in \mathbb{R}$, including $t = 0$, set $\eta_0(t) = 1 - \sum_{k=1}^\infty \eta(2^{-k}t)$. Then $\eta_0 \in \mathcal{C}_0^\infty(\mathbb{R})$ is supported in $[-1, +1]$, and $\eta_0(t) + \sum_{k=1}^\infty \eta(2^{-k}t) \equiv 1$ for all $t \in \mathbb{R}$. Recall that $\mathbf{x} \to |\mathbf{x}|$ is a smooth homogeneous norm on $\mathbb{R}^N$. Set

$$\psi^k(\mathbf{x}) = \begin{cases} \psi(\mathbf{x}) \, \eta_0(|\mathbf{x}|) & \text{if } k = 0, \\ 2^{kQ} \, \psi(2^k \cdot \mathbf{x}) \, \eta(|\mathbf{x}|) & \text{if } k \geq 1. \end{cases}$$

Then $2^{-kQ} \psi^k(2^{-k} \cdot \mathbf{x}) = \psi(\mathbf{x}) \eta(2^{-k}|\mathbf{x}|)$ for $k \geq 1$, and so $\psi(\mathbf{x}) = \sum_{k=0}^\infty 2^{-kQ} \psi^k(2^{-k} \cdot \mathbf{x})$. From the choice of η and the definition of η_0 it follows that each ψ^k is supported on the set where $|\mathbf{x}| \leq 1$, and this gives the assertion (a).

Since $\psi \in \mathcal{S}(\mathbb{R}^N)$, it follows that for every $M \in \mathbb{N}$ and every $\alpha \in \mathbb{N}^N$ with $|\alpha| \leq M$ we have

(6.3) $$|\partial^\alpha \psi(\mathbf{x})| \leq ||\psi||_{[M]} \, (1 + |\mathbf{x}|_e)^{-M},$$

where $|\mathbf{x}|_e$ is the Euclidean length of the vector $\mathbf{x} \in \mathbb{R}^N$. If $\mathbf{x}$ belongs to the support of ψ^k when $k \geq 1$, we have $|\mathbf{x}| \geq 2^{-3}$, and so $|2^k \cdot \mathbf{x}| \geq 2^{k-3}$. Assertion (b) then follows easily from equation (6.3) and the fact that $|\mathbf{x}|^{d_1} \lesssim |\mathbf{x}|_e$ if $|\mathbf{x}|_e \geq 1$.

Finally, suppose that ψ has strong cancellation, so that $\int_{\mathbb{R}^{a_k}} \psi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \, d\mathbf{x}_k = 0$ for $1 \leq k \leq n$. It follows from Lemma 5.1 that we can write ψ as a finite sum of terms of the form $\partial_{j_1} \cdots \partial_{j_n} \psi_{j_1, \ldots, j_n}$ where x_{j_k} is a coordinate in $\mathbf{x}_k$ for $1 \leq k \leq n$, and each function $\psi_{j_1, \ldots, j_n} \in \mathcal{S}(\mathbb{R}^N)$ is normalized relative to ψ. Using assertions (a) and (b), we can write

$$\psi_{j_1, \ldots, j_n}(\mathbf{x}) = \sum_{k=0}^\infty 2^{-kQ} \psi_{j_1, \ldots, j_n}^k (2^{-k} \cdot \mathbf{x}),$$

and so

$$\partial_{j_1} \cdots \partial_{j_n} \psi_{j_1,\dots,j_n}(\mathbf{x}) = \sum_{k=0}^{\infty} 2^{-k(d_{j_1}+\cdots+d_{j_n}+Q)} \partial_{j_1} \cdots \partial_{j_n} \psi_{j_1,\dots,j_n}^k (2^{-k} \cdot \mathbf{x}).$$

Since each term $2^{-k(d_{j_1}+\cdots+d_{j_n})} \partial_{j_1} \cdots \partial_{j_n} \psi_{j_1,\dots,j_n}^k$ has strong cancellation, summing over a finite number of such terms establishes (c), and completes the proof. $\qquad\square$

Finally, we will need the following result, which provides a decomposition of test functions in $\mathcal{C}_0^\infty(\mathbb{R}^N)$:

Lemma 6.6. *Let $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ have compact support in the unit ball, and suppose that φ has cancellation in $\mathbf{x}_1$, i.e., $\int_{\mathbb{R}^{a_1}} \varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \, d\mathbf{x}_1 = 0$. Then there are functions $\{\varphi^j\} \subseteq \mathcal{C}_0^\infty(\mathbb{R}^N)$ such that*

$$\varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{j=-\infty}^{0} 2^{-kQ_1} \varphi^j(2^{-j} \cdot \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n),$$

where Q_1 is the homogeneous dimension of $\mathbb{R}^{a_1}$. Moreover, these functions have the following properties:

(a) *each φ^j is supported in the unit ball;*

(b) *each φ^j is normalized relative to the function φ;*

(c) *for $-\infty < j \leq 0$ we have $\varphi^j(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = 0$ if $|\mathbf{x}_1| \leq \frac{1}{8}$;*

(d) *if φ has strong cancellation, then each function φ^j has strong cancellation.*

Proof. Choose $\eta \in \mathcal{C}_0^\infty(\mathbb{R})$ such that $\eta(t)$ vanishes if $t \geq 2$ or $t \leq \frac{1}{4}$, and such that $\sum_{l \leq 0} \eta(2^{-l}t) = 1$ for $0 < t \leq 1$. Put $A_1(\mathbf{x}_2, \dots, \mathbf{x}_n) = 0$, and for $j \leq 0$, put

$$\chi_j(\mathbf{x}_1) = \eta\left(2^{-j}|\mathbf{x}_1|\right)$$

$$\widetilde{\chi}_j(\mathbf{x}_1) = \chi_j(\mathbf{x}_1)\left[\int_{\mathbb{R}^{a_1}} \chi_j(\mathbf{x}_1) \, d\mathbf{x}_1\right]^{-1}$$

$$a_j(\mathbf{x}_2, \dots, \mathbf{x}_n) = \int_{\mathbb{R}^{a_1}} \varphi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)\chi_j(\mathbf{x}_1) \, d\mathbf{x}_1,$$

$$A_j(\mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{s=j}^{0} a_s(\mathbf{x}_2, \dots, \mathbf{x}_n).$$

We have $\sum_{j \leq 0} \chi_j(\mathbf{x}_1) \equiv 1$ for $0 < |\mathbf{x}_1| \leq 1$, and since φ is supported in the unit ball, we can sum by parts for $\mathbf{x}_1 \neq 0$ to get

$$\varphi(\mathbf{x}) = \sum_{j \leq 0} \varphi(\mathbf{x}_1, \dots, \mathbf{x}_n)\chi_j(\mathbf{x}_1)$$

$$= \sum_{j \leq 0} \left[\varphi(\mathbf{x}_1, \dots, \mathbf{x}_n)\chi_j(\mathbf{x}_1) - \widetilde{\chi}_j(\mathbf{x}_1)a_j(\mathbf{x}_2, \dots, \mathbf{x}_n)\right]$$

$$+ \sum_{j \leq 0}\left(\widetilde{\chi}_j(\mathbf{x}_1) - \widetilde{\chi}_{j-1}(\mathbf{x}_1)\right)A_j(\mathbf{x}_2, \dots, \mathbf{x}_n) = \sum_{j \leq 0} \widetilde{\varphi}_j(\mathbf{x}).$$

Let

$$\varphi^j(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = 2^{jQ_1} \widetilde{\varphi}_j(2^j \cdot \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n).$$

Then

$$(6.4) \quad \varphi^j(\mathbf{x}) = 2^{jQ_1} \big[\varphi(2^j \cdot \mathbf{x}_1, \ldots, \mathbf{x}_n) \chi_0(\mathbf{x}_1) - 2^{-jQ_1} \widetilde{\chi}_0(\mathbf{x}_1) \, a_j(\mathbf{x}_2, \ldots, \mathbf{x}_n) \\ + 2^{jQ_1} \big(\widetilde{\chi}_0(\mathbf{x}_1) - 2^{-Q_1} \widetilde{\chi}_0 \, (2 \cdot \mathbf{x}_1) \big) \, A_j(\mathbf{x}_2, \ldots, \mathbf{x}_n) \big],$$

and

$$(6.5) \qquad \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{j=-\infty}^{0} 2^{-jQ_1} \, \varphi^j(2^{-j} \cdot \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n).$$

The functions $\{a_j\}$ and hence also the functions $\{A_j\}$ are infinitely differentiable functions supported in the unit ball of $\mathbb{R}^{a_2} \oplus \cdots \oplus \mathbb{R}^{a_n}$. Moreover, it follows from the fact that $\chi_j(\mathbf{x}_1)$ is supported on the set $|\mathbf{x}_1| \leq 2^{j+1}$ that there is a constant C so that for each integer m, we have

$$(6.6) \qquad\qquad ||a_j||_{(m)} \leq C \, 2^{jQ_1} ||\varphi||_{(m)}.$$

The function $\sum_{s=-\infty}^{j} \chi_s(\mathbf{x}_1)$ is also supported on the set $|\mathbf{x}_1| \leq 2^{j+1}$, and is bounded independently of j. We have

$$A_j(\mathbf{x}_2, \ldots, \mathbf{x}_n) + \int_{\mathbb{R}^{a_1}} \varphi(\mathbf{x}_1) \sum_{s=-\infty}^{j+1} \chi_s(\mathbf{x}_1) \, d\mathbf{x}_1 = \int_{\mathbb{R}^{a_1}} \varphi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \, d\mathbf{x}_1 = 0,$$

and it thus follows that we also have

$$(6.7) \qquad\qquad ||A_j||_{(m)} \leq C \, 2^{jQ_1} ||\varphi||_{(m)}.$$

It is clear from our construction that each function φ^j has compact support in the unit ball. It follows from equation (6.4) that each φ^j vanishes when $|\mathbf{x}_1| \leq \frac{1}{8}$, and also that if φ has strong cancellation, then $\int_{\mathbb{R}^{a_k}} \varphi^j(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \, d\mathbf{x}_k = 0$ for $1 \leq k \leq n$. Finally, equations (6.4), (6.6), and (6.7) show that $||\varphi^j||_{(m)} \leq C \, 2^{-jQ_1} ||\varphi||_{(m)}$. This completes the proof. $\qquad \square$

Let $\mathcal{K}$ be a flag kernel adapted to the flag $\mathcal{F}$, and let $m = \widehat{\mathcal{K}}$ be the flag multiplier on the flag $\mathcal{F}^*$ which is the Fourier transform of $\mathcal{K}$. Choose a function $\eta \in \mathcal{C}_0^\infty(\mathbb{R})$ supported in $[\frac{1}{2}, 4]$ such that $\sum_{j \in \mathbb{Z}} \eta(2^j t) \equiv 1$ for all $t > 0$. For each $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$, set

$$(6.8) \qquad\qquad \eta_I(\xi) = \eta(2^{i_1}|\xi_1|) \cdots \eta(2^{i_n}|\xi_n|).$$

Note that η_I is supported where $|\xi_j| \approx 2^{-i_j}$ for $1 \leq j \leq n$. We shall establish the following result:

Lemma 6.7. *Let m be a flag multiplier relative to the flag $\mathcal{F}^*$ given in (6.2). Then*

$$(6.9) \qquad\qquad m(\xi) = \sum_{I \in \mathcal{E}_n} m_0(\xi) \, \eta_I(\xi) + \sum_{k=1}^{n} m_k(\xi)$$

where m_0 is the Fourier transform of a flag kernel relative to the flag $\mathcal{F}$, and for $1 \leq k \leq n$, the function m_k is the Fourier transform of a flag kernel adapted to a flag strictly coarser *than $\mathcal{F}$.*

Proof. Let θ be a smooth function on $\mathbb{R}$ supported where $t \geq 10$ such that $\theta(t) = 1$ for $t \geq 20$. Write

$$m(\xi) = m(\xi)\theta\left(|\xi_{n-1}|\,|\xi_n|^{-1}\right) + m(\xi)\left[1 - \theta\left(|\xi_{n-1}|\,|\xi_n|^{-1}\right)\right] = n_1(\xi) + m_1(\xi).$$

On the support of $\theta'\left(|\xi_{n-1}|\,|\xi_n|^{-1}\right)$ we have $|\xi_{n-1}| \sim |\xi_n|$. Also, by homogeneity we have

$$\left|\partial_{\xi_{n-1}}^{\alpha_{n-1}}(|\xi_{n-1}|)\right| \leq C|\xi_{n-1}|^{1-|\alpha_{n-1}|} \quad \text{and} \quad \left|\partial_{\xi_n}^{\alpha_n}(|\xi_n|)\right| \leq C|\xi_n|^{1-|\alpha_n|}.$$

Thus Lemma 6.4 implies that $n_1(\xi)$ and $m_1(\xi)$ are flag multipliers relative to the flag (6.2). On the support of m_1 we have $|\xi_{n-1}|\,|\xi_n|^{-1} \leq 20$. Thus we can group together the variables ξ_n and ξ_{n-1}, and it follows that $m_1(\xi)$ is a flag kernel relative to a flag *coarser* than $\mathcal{F}^*$. Also $|\xi_{n-1}| \geq 10|\xi_n|$ on the support of n_1.

Next write

$$n_1(\xi) = n_1(\xi)\theta\left(|\xi_{n-2}|\,|\xi_{n-1}|^{-1}\right) + n_1(\xi)\left[1 - \theta\left(|\xi_{n-2}|\,|\xi_{n-1}|^{-1}\right)\right] = n_2(\xi) + m_2(\xi).$$

Since $\theta'\left(|\xi_{n-2}|\,|\xi_{n-1}|^{-1}\right)$ is supported where $|\xi_{n-2}| \sim |\xi_{n-1}|$ and $|\xi_{n-1}| \geq 10|\xi_n|$, it again follows that n_2 and m_2 are Fourier transforms of flag kernels relative to the flag $\mathcal{F}^*$. On the support of m_2 we have $|\xi_{n-2}|\,|\xi_{n-1}|^{-1} \leq 20$ and $|\xi_{n-1}|\,|\xi_n|^{-1} \leq 20$. Thus we can group together the variables ξ_n, ξ_{n-1} and ξ_{n-2}, and it follows that $m_2(\xi)$ is a flag kernel relative to a flag *coarser* than $\mathcal{F}^*$. Also $|\xi_{n-1}| \geq 10|\xi_{n-1}|$, and $|\xi_{n-1}| \geq 10|\xi_n|$ on the support of n_2.

We proceed inductively to see that

$$m(\xi) = m_0(\xi) + \sum_{k=1}^{n} m_k(\xi)$$

where each $m_k(\xi)$ is the Fourier transform of a coarser flag kernel and $m_0(\xi) = n_n(\xi)$ is supported where $|\xi_j| \geq 10|\xi_{j+1}|$ for $1 \leq j \leq n-1$.

From our choice of η, it follows that

$$
\begin{aligned}
1 &= \sum_{J=(j_1,\ldots,j_n)\in\mathbb{Z}^n} \eta\left(2^{j_1}|\xi_1|\right) \cdots \eta\left(2^{j_n}|\xi_n|\right) \\
&= \sum_{J\in\mathcal{E}_n} \eta\left(2^{j_1}|\xi_1|\right) \cdots \eta\left(2^{j_n}|\xi_n|\right) + \sum_{J\in\mathbb{Z}^n-\mathcal{E}_n} \eta\left(2^{j_1}|\xi_1|\right) \cdots \eta\left(2^{j_n}|\xi_n|\right)
\end{aligned}
$$

Thus if $\sum_{J\in\mathcal{E}_n} \eta\left(2^{j_1}|\xi_1|\right) \cdots \eta\left(2^{j_n}|\xi_n|\right) \neq 1$, there is an n-tuple $J = (j_1,\ldots,j_n)$ and integers $1 \leq r < s \leq n$ such that $j_r > j_s$ and $\eta\left(2^{j_r}|\xi_r|\right) \eta\left(2^{j_s}|\xi_s|\right) \neq 0$. Since η is supported on $[\frac{1}{2}, 4]$, it follows that $|\xi_r| \leq 4\,2^{-j_r} \leq 4\,2^{-j_s} \leq 8|\xi_s|$. However, on the support of m_0 we have $|\xi_r| > 10|\xi_s|$. Thus on the support of m_0 we have

$$1 = \sum_{J\in\mathcal{E}_n} \eta\left(2^{j_1}|\xi_1|\right) \cdots \eta\left(2^{j_n}|\xi_n|\right),$$

and so

$$m(\xi) = \sum_{J \in \mathcal{E}_n} m_0(\xi) \eta\left(2^{j_1} |\xi_1|\right) \cdots \eta\left(2^{j_n} |\xi_n|\right) + \sum_{k=1}^{n} m_k(\xi),$$

which completes the proof.   □

We now turn to the proof of Theorem 6.1. If we write $\mathcal{K}_j$ as the inverse Fourier transform of the flag multiplier m_j of Lemma 6.7, we have shown that

$$\mathcal{K} = \mathcal{K}_0 + \sum_{j=1}^{n} \mathcal{K}_j,$$

and for $1 \le j \le n$, $\mathcal{K}_j$ is a flag kernel adapted to a flag which is coarser than $\mathcal{F}$. Also since $m_0(\xi) = \sum_{J \in \mathcal{E}_n} m_0(\xi) \eta_I(\xi)$, we can write

$$(6.10) \qquad \mathcal{K}_0 = \sum_{J \in \mathcal{E}_n} \left[\Psi^I\right]_I$$

where

$$\widehat{\Psi^I}(\xi) = m_0(2^{-i_1} \cdot \xi_1, \ldots, 2^{-i_n} \cdot \xi_n) \cdot \eta(|\xi_1|) \cdots \eta(|\xi_n|),$$

and the sum converges in the sense of distributions. The differential inequalities for m_0 imply that each function $\Psi^I \in \mathcal{S}(\mathbb{R}^N)$, with Schwartz norms uniformly bounded in I. Also since Ψ^I vanishes on the coordinate axes, for $1 \le k \le n$ we have

$$(6.11) \qquad \int \Psi^I(\mathbf{x}_1, \ldots, \mathbf{x}_n) \, d\mathbf{x}_k = 0.$$

In order to complete the proof of Theorem 6.1, it remains to show that we can replace the Schwartz functions Ψ^I in equation (6.10) by functions in $\mathcal{C}_0^\infty(\mathbb{R}^N)$, all supported in the unit ball with strong cancellation, and which vanish when $|\mathbf{x}_1| < \epsilon$. This is done in two steps, using Lemmas 6.5 and 6.6.

First, according to Lemma 6.5, for each $I \in \mathcal{E}_n$ there are functions $\psi^{k,I} \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, each supported in the unit ball and having strong cancellation, so that

$$\psi^I(\mathbf{x}) = \sum_{k=0}^{\infty} 2^{-kQ} \psi^{k,I}(2^{-k} \cdot \mathbf{x}),$$

and for every $\Delta > 0$ and every positive integer m there is an integer p_m so that

$$||\psi^{k,I}||_{(m)} \le 2^{-k\Delta} \, ||\psi^I||_{[m+p_m]}.$$

Thus

$$[\psi^I]_I(\mathbf{x}) = \sum_{k=0}^{\infty} 2^{-Q_1(k+i_1) - \cdots - Q_n(k+i_n)} \, \psi^{k,I}\left(2^{-(i_1+k)} \cdot \mathbf{x}_1, \ldots, 2^{-(i_n+k)} \cdot \mathbf{x}_n\right),$$

and so formally

$$
\sum_{I \in \mathcal{E}_n} [\psi^I]_I(\mathbf{x}) = \sum_{k=0}^{\infty} \sum_{I \in \mathcal{E}_n} 2^{-Q_1(k+i_1)-\cdots-Q_n(k+i_n)} \psi^{k,I}\big(2^{-(i_1+k)} \cdot \mathbf{x}_1, \ldots, 2^{-(i_n+k)} \cdot \mathbf{x}_n\big)
$$

$$
= \sum_{k=0}^{\infty} \sum_{J \in \mathcal{E}_n} 2^{-Q_1(j_1)-\cdots-Q_n(j_n)} \psi^{k,I}\big(2^{-j_1} \cdot \mathbf{x}_1, \ldots, 2^{-j_n} \cdot \mathbf{x}_n\big)
$$

$$
= \sum_{J \in \mathcal{E}_n} 2^{-Q_1(j_1)-\cdots-Q_n(j_n)} \Big[\sum_{k=0}^{\infty} \psi^{k,I}\Big]\big(2^{-j_1} \cdot \mathbf{x}_1, \ldots, 2^{-j_n} \cdot \mathbf{x}_n\big)
$$

$$
= \sum_{J \in \mathcal{E}_n} \Big[\sum_{k=0}^{\infty} \psi^{k,I}\Big]_J(\mathbf{x}).
$$

The estimates we have on the functions $\{\psi^{k,I}\}$ show that the series $\sum_{k=0}^{\infty} \psi^{k,I} = \widetilde{\varphi}^I$ converges in $\mathcal{C}_0^{\infty}(\mathbb{R}^N)$ to a function supported in the unit ball which has strong cancellation. This formal calculation is easily justified by applying it to finite sums of dilates. Thus we have shown $\mathcal{K}_0 = \sum_{I \in \mathcal{E}_n} [\widetilde{\varphi}^I]_I$, with convergence in the sense of distributions, where the functions $\widetilde{\varphi}^I \in \mathcal{C}_0^{\infty}(\mathbb{R}^N)$ all have support in the unit ball, have strong cancellation, and are uniformly bounded in each semi-norm $\| \cdot \|_{(m)}$.

Finally, Lemma 6.6 shows that for each $I \in \mathcal{E}_n$, there exist functions $\varphi^{j,I} \in \mathcal{C}_0^{\infty}(\mathbb{R}^N)$ with strong cancellation, each supported in the unit ball, normalized relative to $\widetilde{\varphi}^I$, and vanishing when $|\mathbf{x}_1| \leq \frac{1}{8}$ so that

$$
\widetilde{\varphi}^I(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \sum_{j=-\infty}^{0} 2^{-jQ_1} \varphi^{j,I}\big(2^{-j} \cdot \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\big).
$$

We then have

$$
\big[\widetilde{\varphi}^I\big]_I(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)
$$

$$
= \sum_{j=-\infty}^{0} 2^{-(j+i_1)Q_1-i_2Q_2-\cdots-i_nQ_n} \varphi^{j,I}\big(2^{-(j+i_1)} \cdot \mathbf{x}_1, 2^{-i_2} \cdot \mathbf{x}_2, \ldots, 2^{-i_n} \cdot \mathbf{x}_n\big),
$$

and so, since we are summing only over non-positive indices j, we have

$$
\sum_{I \in \mathcal{E}_n} [\widetilde{\varphi}^I]_I(\mathbf{x}) = \sum_{j=-\infty}^{0} \sum_{I \in \mathcal{E}_n} 2^{-(j+i_1)Q_1-i_2Q_2-\cdots-i_nQ_n} \varphi^{j,I}\big(2^{-(j+i_1)} \cdot \mathbf{x}_1, \ldots, 2^{-i_n} \cdot \mathbf{x}_n\big)
$$

$$
= \sum_{j=-\infty}^{0} \sum_{I \in \mathcal{E}_n} 2^{-i_1Q_1-i_2Q_2-\cdots-i_nQ_n} \varphi^{j,I}\big(2^{-i_1} \cdot \mathbf{x}_1, \ldots, 2^{-i_n} \cdot \mathbf{x}_n\big)
$$

$$
= \sum_{I \in \mathcal{E}_n} 2^{-i_1Q_1-i_2Q_2-\cdots-i_nQ_n} \Big[\sum_{j=-\infty}^{0} \varphi^{j,I}\Big]\big(2^{-i_1} \cdot \mathbf{x}_1, \ldots, 2^{-i_n} \cdot \mathbf{x}_n\big)
$$

$$
= \sum_{I \in \mathcal{E}_n} \Big[\sum_{j=-\infty}^{0} \varphi^{j,I}\Big]_I(\mathbf{x}).
$$

The estimates we have on the functions $\{\varphi^{j,I}\}$ show that the series $\sum_{j=-\infty}^{0} \varphi^{j,I} = \varphi^{I}$ converges in $\mathcal{C}_0^\infty(\mathbb{R}^N)$ to a function supported in the unit ball which has strong cancellation, and which vanishes when $|\mathbf{x}_1| \leq \frac{1}{8}$. This at last completes the proof of Theorem 6.1.

6.2. Dyadic sums with weak cancellation

The main result of this section, Theorem 6.8, is a strengthening of the converse to Theorem 6.1 which would assert that sums of dilates of appropriate compactly supported bump functions with strong cancellation are flag kernels; we consider sums of Schwartz functions instead of compactly supported functions, and more critically, we assume only weak cancellation instead of strong cancellation relative to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$.

Theorem 6.8. *For each $I \in \mathcal{E}_n$, let $\varphi^I \in \mathcal{S}(\mathbb{R}^N)$, and suppose*

(a) *there are constants $C_N > 0$ so that $||\varphi^I||_{[N]} \leq C_N$ for each $I \in \mathcal{E}_n$ and for each $N \geq 0$;*

(b) *there is a constant $\epsilon > 0$ so that each φ^I has weak cancellation with respect to I with parameter ϵ relative to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$.*

Then we have the following conclusions:

(1) *For any finite set $F \subset \mathcal{E}_n$, the function $K_F = \sum_{I \in F} [\varphi^I]_I \in \mathcal{S}(\mathbb{R}^N)$ defines a flag kernel $\mathcal{K}_F$ for the flag $\mathcal{F}$ with bounds which are independent of the set F.*

(2) *Let $F_1 \subset F_2 \subset \cdots \subset F_m \subset \cdots$ be any increasing sequence of finite subsets of E_N with $E_N = \bigcup_{m=1}^{\infty} F_m$. Then for any test function $\psi \in \mathcal{S}(\mathbb{R}^N)$,*

$$\lim_{m \to \infty} \langle K_{F_m}, \psi \rangle = \lim_{m \to \infty} \sum_{I \in F_m} \int_{\mathbb{R}^N} [\varphi^I]_I(\mathbf{x}) \psi(\mathbf{x}) \, dx$$

exists and defines a flag kernel $\mathcal{K} \in \mathcal{S}'(\mathbb{R}^N)$ which is independent of the choice of the finite subsets. We write this limit as

$$\mathcal{K} = \lim_{F \nearrow E_N} \sum_{I \in F} [\varphi^I]_I.$$

Since the proof of this result is somewhat involved, let us indicate the main steps. If $F \subset \mathcal{E}_n$ is any finite set, let $\mathcal{K}_F = \sum_{I \in F} [\varphi^I]_I$. We show (in Proposition 6.9) that $K_F = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$ satisfies the correct size estimates, and also (in Proposition 6.11) that K_F satisfies the cancellation conditions with constants independent of F of Definition 2.3. This will establish part (1) of Theorem 6.8. To establish the existence of the limit in part (2), we use the weak cancellation of the functions $\{\varphi^I\}$ to show that for any test function ψ, the bracket $\langle \mathcal{K}_F, \psi \rangle$ can be rewritten as the integral of ψ and its derivatives against locally integrable functions. The existence of the limit then follows from the Lebesgue dominated convergence theorem.

We now turn to the proofs. Note that in the next proposition we impose no cancellation conditions on the functions $\{\varphi^I\}$.

Proposition 6.9. *For each $I \in \mathcal{E}_n$, let $\varphi^I \in \mathcal{S}(\mathbb{R}^N)$, and suppose there are constants C_M so that for all $I \in \mathcal{E}_n$ and all M, $\|\varphi^I\|_{[M]} \leq C_M$. Let $F \subset \mathcal{E}_n$ be a finite subset, and let $K_F(\mathbf{x}) = \sum_{I \in F}[\varphi^I]_I(\mathbf{x})$. For any $\alpha = (\bar{\alpha}_1, \ldots, \bar{\alpha}_n) \in \mathbb{N}^{a_1} \times \cdots \times \mathbb{N}^{a_n}$ with $|\alpha| \leq M$, there is a constant A_M independent of the finite set F so that*

$$\left| \partial_{\mathbf{x}}^\alpha K_F(\mathbf{x}) \right| \leq A_M \prod_{j=1}^n \left[N_1(\mathbf{x}_1) + N_2(\mathbf{x}_2) + \cdots + N_j(\mathbf{x}_j) \right]^{-(Q_j + \llbracket \bar{\alpha}_j \rrbracket)}.$$

Proof. It follows from Proposition 4.3 that

$$\left| \partial_{\mathbf{x}}^\alpha K_F(\mathbf{x}) \right| = \left| \sum_{I \in F} 2^{-\sum_{j=1}^n i_j (Q_j + \llbracket \bar{\alpha}_j \rrbracket)} \partial_{\mathbf{x}}^{\bar{\alpha}}[\varphi^I](2^{-i_1} \cdot \mathbf{x}_1, \ldots, 2^{-i_n} \cdot \mathbf{x}_n) \right|.$$

Since $\varphi^I \in \mathcal{S}(\mathbb{R}^n)$, for any M we have the estimate

$$\left| \partial_{\mathbf{x}}^{\bar{\alpha}}[\varphi^I](2^{-i_1}\mathbf{x}_1, \ldots, 2^{-i_n}\mathbf{x}_n) \right| \leq C_M \left(1 + \sum_{k=1}^N 2^{-i_k} N_k(\mathbf{x}_k) \right)^{-M}.$$

Thus

$$\left| \partial_{\mathbf{x}}^{\bar{\alpha}} K_F(\mathbf{x}) \right| \leq C_M \sum_{I \in F} 2^{-\sum_{j=1}^n i_k (Q_k + \llbracket \bar{\alpha}_k \rrbracket)} \left(1 + \sum_{k=1}^N 2^{-i_k} N_k(\mathbf{x}_k) \right)^{-M}.$$

Proposition 6.9 thus follows from estimate (10.2) in Prop. 10.1 in Appendix II. □

The next result provides estimates that will be used in establishing the cancellation conditions (part (b) of Definition 2.3) for finite sums $\mathcal{K}_F = \sum_{I \in F}[\varphi^I]_I$. Recall that these cancellation conditions involve integrals of the form

$$\int K(\mathbf{x}_1, \ldots, \mathbf{x}_n)\, \psi(R_1 \cdot \mathbf{x}_{m_1}, \ldots, R_\beta \cdot \mathbf{x}_{m_\beta})\, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_\beta}$$

where $\{m_1, \ldots, m_\beta\}$ is a non-empty subset of $\{1, \ldots, n\}$. Let $M = \{m_1, \ldots, m_\beta\}$, and let $L = \{l_1, \ldots, l_\alpha\} = \{1, \ldots, n\} \setminus M$. Let $N_\alpha = a_{l_1} + \cdots + a_{l_\alpha}$ and $N_\beta = a_{m_1} + \cdots + a_{m_\beta}$ so that $N_\alpha + N_\beta = N$. For $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^N$, write

$$\mathbf{x} = (\mathbf{x}', \mathbf{x}'') \quad \text{with} \quad \begin{cases} \mathbf{x}' = (\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_\alpha}) \in \mathbb{R}^{N_\alpha} \\ \mathbf{x}'' = (\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}) \in \mathbb{R}^{N_\beta} \end{cases}.$$

For $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$, write

$$I = (I', I'') \quad \text{where} \quad \begin{cases} I' = (i_{l_1}, \ldots, i_{l_\alpha}) \in E_\alpha \\ I'' = (i_{m_1}, \ldots, i_{m_\beta}) \in E_\beta \end{cases}.$$

If $R = (R_1, \ldots, R_\beta)$ is a β-tuple of positive real numbers, write

$$R \cdot \mathbf{x}'' = (R_1 \cdot \mathbf{x}_{m_1}, \ldots, R_\beta \cdot \mathbf{x}_{m_\beta}),$$

$$2^{I''} R \cdot \mathbf{x}'' = (2^{i_{m_1}} R_1 \cdot \mathbf{x}_{m_1}, \ldots, 2^{i_{m_\beta}} R_\beta \cdot \mathbf{x}_{m_\beta}).$$

Finally, let $P(M)$ denote the set of all partitions of the set $M = \{m_1, \ldots, m_\beta\}$ into two disjoint (possibly empty) subsets A and B.

Proposition 6.10. *With the above notation, let $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$, and let $\varphi \in \mathcal{S}(\mathbb{R}^N)$. Let $\psi \in \mathcal{C}_0^\infty (\mathbb{R}^{N_\beta})$ so that ψ can be regarded as a function of the variables $\mathbf{x}'' = (\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta})$. Let $R = (R_1, \ldots, R_\beta)$ be a β-tuple of positive real numbers. Define Φ on $\mathbb{R}^{N_\alpha}$ by setting*

$$\Phi(\mathbf{x}') = \int_{\mathbb{R}^{N_\beta}} [\varphi]_I(\mathbf{x}', \mathbf{x}'') \, \psi(R_1 \cdot \mathbf{x}_{m_1}, \ldots, R_\beta \cdot \mathbf{x}_{m_\beta}) \, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_\beta}.$$

(1) *There exists $\Theta \in \mathcal{S}(\mathbb{R}^{N_\alpha})$, normalized relative to φ and ψ (with constants independent of $(R_1, \ldots, R_\beta)$) such that $\Phi(\mathbf{x}') = [\Theta]_{I'}(\mathbf{x}')$.*

(2) *If φ has weak cancellation with respect to I, there are constants C and ϵ independent of $R = (R_1, \ldots, R_\beta)$ so that $\Theta = \sum_{(A,B)\in P(M)} \Theta_{A,B}$, where for each partition $M = A \cup B$ and $n \notin A$,*

$$|\Theta_{A,B}(\mathbf{x}')| \leq C \prod_{m_r \in A} 2^{-\epsilon(i_{m_r+1} - i_{m_r})} \prod_{m_s \in B} \min \left\{ (R_s 2^{i_{m_s}})^{+\epsilon}, (R_s 2^{i_{m_s}})^{-\epsilon} \right\}.$$

Thus for each partition $M = A \cup B$ of the set of variables $\{\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}\}$, $|\Theta_{A,B}(\mathbf{x}')|$ is small due to two kinds of gains: there is a gain $2^{-\epsilon(i_{m_r+1} - i_{m_r})}$ for every index $m_r \in A$, and there is a gain $\min \left\{ (R_r 2^{i_{m_r}})^{+\epsilon}, (R_r 2^{i_{m_r}})^{-\epsilon} \right\}$ for every index $m_r \in B$.

Proof. Make the change of variables

$$\mathbf{x}'' = (\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}) \mapsto (2^{i_{m_1}} \cdot \mathbf{x}_{m_1}, \ldots, 2^{i_{m_\beta}} \cdot \mathbf{x}_{m_\beta}).$$

Then

$$\Phi(\mathbf{x}') = 2^{-\sum_{k=1}^{n} i_k Q_k} \int_{\mathbb{R}^{N_\beta}} \varphi(2^{-i_1}\mathbf{x}_1, \ldots, 2^{-i_n}\cdot\mathbf{x}_n)\psi(R_1\cdot\mathbf{x}_{m_1}, \ldots, R_\beta\cdot\mathbf{x}_{m_\beta}) \, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_\beta}$$

$$= \int_{\mathbb{R}^{N_\beta}} [\varphi]_{I'}(\mathbf{x}', \mathbf{x}'')\psi(R_1 2^{i_{m_1}} \cdot \mathbf{x}_{m_1}, \ldots, R_\beta 2^{i_{m_\beta}} \cdot \mathbf{x}_{m_\beta}) \, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_\beta}$$

$$= [\Theta]_{I'}(\mathbf{x}'),$$

where

$$(6.12) \quad \Theta(\mathbf{x}') = \int_{\mathbb{R}^{N_\beta}} \varphi(\mathbf{x}', \mathbf{x}'') \, \psi(R_1 2^{i_{m_1}} \cdot \mathbf{x}_{m_1}, \ldots, R_\beta 2^{i_{m_\beta}} \cdot \mathbf{x}_{m_\beta}) \, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_\beta}.$$

Now

$$(1 + |\mathbf{x}'|)^M \partial_{\mathbf{x}'}^{\bar{\alpha}} \Theta(\mathbf{x}') = \int_{\mathbb{R}^{N_\beta}} (1 + |\mathbf{x}'|)^M \partial_{\mathbf{x}'}^{\bar{\alpha}} \varphi(\mathbf{x}', \mathbf{x}'') \, \psi(2^{I''} R \cdot \mathbf{x}'') \, d\mathbf{x}'',$$

and we can estimate the $\mathcal{S}(\mathbb{R}^{N_\alpha})$-seminorms of Θ in terms of those of φ and the supremum of $|\psi|$, independent of the choice of R. This establishes assertion (1).

To prove the decomposition and additional size estimates for Θ asserted in part (2), assume that φ has weak cancellation relative to I. Since the definiton (6.12) of Θ involves integration with respect to the variables $\{\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta}\}$, we will only use the weak cancellation in these variables. We can use the first of the Remarks 5.6 to write φ as a sum of terms of the form

$$\Big(\prod_{s \in A} 2^{-\epsilon(i_{s+1} - i_s)} \Big) \Big(\prod_{t \in B} \partial_{\sigma(t)} \Big) [\widetilde{\varphi}_{A,B,\sigma}]$$

where $\{m_1, \ldots, m_\beta\} = A \cup B$ with $A \cap B = \emptyset$ and $n \in B$ if $n \in \{m_1, \ldots, m_\beta\}$. Also, $\sigma \colon B \to \{1, \ldots, N\}$ so that $\sigma(m_\ell) \in J_{m_\ell}$, and each $\widetilde{\varphi}_{A,B,\sigma}$ is normalized relative to φ.

If we write $B = \{m_{\ell_1}, \ldots, m_{\ell_s}\}$, it follows that $\Theta(\mathbf{x}')$ is a finite sum of terms of the form

$$I(A; B; \sigma) = \prod_{s \in A} 2^{-\epsilon(i_{s+1} - i_s)} \int_{\mathbb{R}^{N_\beta}} \partial_{\sigma(m_{\ell_1})} \cdots \partial_{\sigma(m_{\ell_s})} [\widetilde{\varphi}_{A;B;\sigma}]$$

$$\cdot \psi(R_1 2^{i_{m_1}} \cdot \mathbf{x}_{m_1}, \ldots, R_\beta 2^{i_{m_\beta}} \cdot \mathbf{x}_{m_\beta}) \, d\mathbf{x}_{m_1} \cdots d\mathbf{x}_{m_\beta}.$$

We can use integration by parts to move the differentiations $\partial_{\sigma(m_{\ell_1})} \cdots \partial_{\sigma(m_{\ell_s})}$ from $\widetilde{\varphi}_{A;B;\sigma}$ to ψ. Differentiating the function $\psi(R_1 2^{i_{m_1}} \cdots \mathbf{x}_{m_1}, \ldots, R_\beta 2^{i_{m_\beta}} \cdot \mathbf{x}_{m_\beta})$ with respect to the variable $x_{\sigma(m_{\ell_k})}$ with $\sigma(m_{\ell_k}) \in J_{m_{\ell_k}}$ brings out a factor $(R_{m_{\ell_k}} 2^{i_{m_{\ell_k}}})^{d_{\sigma(m_{\ell_k})}}$, and so we have the estimate

$$\big| I(j_1, \ldots, j_\beta; A; \sigma) \big| \lesssim \prod_{s \in A} 2^{-\epsilon(i_{s+1} - i_s)} \prod_{r \in B} (R_r 2^{i_r})^{d_{\sigma(r)}}.$$

On the other hand, without integrating by parts, since ψ has compact support, the integral in the variables $\{\mathbf{x}_{m_r} \,|\, r \in B\}$ is taken over the set where $|\mathbf{x}_{m_r}| \lesssim (R_r 2^{i_{m_r}})^{Q_{m_r}}$ for $r \in B$, and this set has volume bounded by a constant times $\prod_{r \in B} (R_r 2^{i_{m_r}})^{-Q_{m_r}}$. It follows that there exists $\epsilon > 0$ so that

$$(6.13) \quad \big| I(j_1, \ldots, j_\beta; A; \sigma) \big| \lesssim \prod_{s \in A} 2^{-\epsilon(i_{s+1} - i_s)} \prod_{r \in B} \min \big\{ (R_r 2^{i_{m_r}})^{+\epsilon}, (R_r 2^{i_{m_r}})^{-\epsilon} \big\}.$$

This completes the proof.   □

We now show that if the functions $\{\varphi^I\}$ have weak cancellation, then the sum $K_F(\mathbf{x})$ satisfies the cancellation condition (condition (b)) of Definition 2.3. We use the same notation as in Proposition 6.10. Thus $L = \{l_1, \ldots, l_\alpha\}$ and $M = \{m_1, \ldots, m_\beta\}$ are complementary subsets of $\{1, 2, \ldots, n\}$, and we set $N_\alpha = a_{l_1} + \cdots + a_{l_\alpha}$ and $N_\beta = a_{m_1} + \cdots + a_{m_\beta}$. If $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^N$, we write $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$, with $\mathbf{x}' = (\mathbf{x}_{l_1}, \ldots, \mathbf{x}_{l_\alpha})$ and $\mathbf{x}'' = (\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta})$. If $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$, write $I' = (i_{l_1}, \ldots, i_{l_\alpha}) \in E_\alpha$ and $I'' = (i_{m_1}, \ldots, i_{m_\beta}) \in E_\beta$

Proposition 6.11. *For each $I \in \mathcal{E}_n$ let $\varphi^I \in \mathcal{S}(\mathbb{R}^N)$, and suppose there are constants C_M so that that for all $I \in \mathcal{E}_n$ and all M, $\|\varphi^I\|_{[M]} \leq C_M$. Let $F \subset \mathcal{E}_n$ be a finite subset, and let $K_F(\mathbf{x}) = \sum_{I \in F} [\varphi^I]_I(\mathbf{x})$. Let $\psi \in \mathcal{C}_0^\infty\left(\mathbb{R}^{N_\beta}\right)$ be a bump function in the variables $\mathbf{x}'' = (\mathbf{x}_{m_1}, \ldots, \mathbf{x}_{m_\beta})$. Let $R = (R_1, \ldots, R_\beta)$ be a β-tuple*

of positive real numbers, and let $\bar{\gamma} = (\gamma_{l_1}, \ldots, \gamma_{l_\alpha}) \in \mathbb{N}^{a_{l_1}} \oplus \cdots \oplus \mathbb{N}^{a_{l_\alpha}}$. There exists a constant $C_{\bar{\gamma}}$, independent of R so that

$$\left| \partial_{\mathbf{x}_{l_1}}^{\gamma_{l_1}} \cdots \partial_{\mathbf{x}_{l_\alpha}}^{\gamma_{l_\alpha}} \int_{\mathbb{R}^{N_\beta}} K_F(\mathbf{x}', \mathbf{x}'') \, \psi(R \cdot \mathbf{x}'') \, d\mathbf{x}'' \right|$$

$$\leq C_{\bar{\gamma}} \prod_{p=1}^{\alpha} \left[N_{l_1}(\mathbf{x}_{l_1}) + \cdots N_{l_p}(\mathbf{x}_{l_p}) \right]^{-Q_{l_p} - \llbracket \gamma_{l_p} \rrbracket}.$$

Proof. Recall the definitions of I' and I'' from just before the statement of the proposition. Using Proposition 6.10 to write $\int [\varphi^I]_{I'}(\mathbf{x}', \mathbf{x}'')\psi(R \cdot \mathbf{x}'') \, d\mathbf{x}'' = [\Theta^I]_{I'}(\mathbf{x}')$, we have

$$\int K_F(\mathbf{x}', \mathbf{x}'')\psi(R \cdot \mathbf{x}'') \, d\mathbf{x}'' = \sum_{I \in F \subset \mathcal{E}_n} \int [\varphi^I]_{I'}(\mathbf{x}', \mathbf{x}'')\psi(R \cdot \mathbf{x}'') \, d\mathbf{x}'' = \sum_{I \in F \subset \mathcal{E}_n} [\Theta^I]_{I'}(\mathbf{x}').$$

Each Θ^I is normalized relative to φ. We write the sum over $I \in F$ as an iterated sum as follows. Let

$$E_1 = \left\{ I' = (i_{l_1}, \ldots, i_{l_\alpha}) \in \mathbb{Z}^\alpha \, \middle| \, (i_1, \ldots, i_n) \in F \right\},$$

and for $I' \in E_1$, let

$$E_2(I') = \left\{ I'' = (i_{m_1}, \ldots, i_{m_\beta}) \} \in \mathbb{Z}^\beta \, \middle| \, (i_1, \ldots i_n) \in F \subset \mathcal{E}_n \right\}.$$

If $I \in F$, we write $I = (I', I'')$ with $I' \in E_1$ and $I'' \in E_2(I')$. Then

$$\int K_F(\mathbf{x}', \mathbf{x}'') \, \psi(R \cdot \mathbf{x}'') \, d\mathbf{x} = \sum_{I' \in E_1} \left[\sum_{I'' \in E_2(I')} \Theta^{(I', I'')} \right]_{I'}(\mathbf{x}').$$

We must show that this sum satisfies the differential inequalities for flags on the space $\mathbb{R}^{a_{l_1}} \oplus \cdots \oplus \mathbb{R}^{a_{l_\alpha}}$, with constants independent of the finite set F. This will follow from Proposition 6.9 provided we can show that for each $I' \in E_1$, the sum $\sum_{I'' \in E_2(I')} \Theta^{(I', I'')}$ converges to a normalized Schwartz function. However, this follows from the estimates in part (2) of Proposition 6.10. $\qquad \square$

We now turn to the proof of Theorem 6.8. As already indicated on page 663, just before the statement of Proposition 6.9, part (1) is an immediate consequence of Proposition 6.9 and Proposition 6.11, so we only need to establish part (2). Let $\psi \in \mathcal{S}(\mathbb{R}^n)$, and for each $I \in \mathcal{E}_n$, let $\varphi^I \in \mathcal{S}(\mathbb{R}^N)$ have weak cancellation relative to I. According to the second of the two Remarks 5.6, we can write

$$\varphi^I = \sum_{\ell=1}^{a_1} \partial_\ell [\varphi_\ell^I] + 2^{-\epsilon(i_2 - i_1)} \varphi_0^I,$$

and so

$$\left[\varphi^I\right]_I = \sum_{\ell=1}^{a_1} (2^{i_1 d_\ell} \partial_\ell)[\varphi_\ell^I]_I + 2^{-\epsilon(i_2 - i_1)} \left[\varphi_0^I\right]_I.$$

Then integrating by parts, we have

$$\int_{\mathbb{R}^N} [\varphi^I(\mathbf{x})]_I \psi(\mathbf{x})\, d\mathbf{x} = -\sum_{\ell=1}^{a_1} \int_{\mathbb{R}^N} 2^{i_1 d_\ell} [\varphi_\ell^I]_I(\mathbf{x}) \frac{\partial \psi}{\partial x_\ell}(\mathbf{x})\, d\mathbf{x} + \int_{\mathbb{R}^N} 2^{-\epsilon(i_2 - i_1)} [\varphi_0^I]_I \psi(\mathbf{x})\, d\mathbf{x}.$$

Thus if $F \subset \mathcal{E}_n$ is a finite subset, and $K_F(\mathbf{x}) = \sum_{I \in F} [\varphi^I]_I(\mathbf{x})$, we have

$$\int_{\mathbb{R}^N} K_F(\mathbf{x}) \psi(\mathbf{x})\, d\mathbf{x}$$

$$= -\sum_{\ell=1}^{a_1} \int_{\mathbb{R}^N} \Big[\sum_{I \in F} 2^{i_1 d_\ell} [\varphi_\ell^I]_I(\mathbf{x}) \Big] \frac{\partial \psi}{\partial x_\ell}(\mathbf{x})\, d\mathbf{x} + \int_{\mathbb{R}^N} \Big[\sum_{I \in F} 2^{-\epsilon(i_2 - i_1)} [\varphi_0^I]_I \Big] \psi(\mathbf{x})\, d\mathbf{x}$$

$$= -\sum_{\ell=1}^{a_1} \int_{\mathbb{R}^N} K_F^\ell(\mathbf{x}) \frac{\partial \psi}{\partial x_\ell}(\mathbf{x})\, d\mathbf{x} + \int_{\mathbb{R}^N} K_F^0(\mathbf{x}) \psi(\mathbf{x})\, d\mathbf{x},$$

where

$$K_F^\ell(\mathbf{x}) = \sum_{I \in F} 2^{i_1 d_\ell} [\varphi_\ell^I]_I(\mathbf{x}), \quad 1 \leq \ell \leq a_1,$$

$$K_F^0(\mathbf{x}) = \sum_{I \in F} 2^{-\epsilon(i_2 - i_1)} [\varphi_0^I]_I.$$

If $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^N$, we have

$$\partial^\alpha K_F^\ell(\mathbf{x}) = \sum_{I \in F} 2^{i_1 d_\ell}\, 2^{-\sum_{j=1}^n i_j(Q_j + [\![\alpha_j]\!])}\, \partial^\alpha [\varphi_\ell^I](2^{-I} \cdot \mathbf{x}),$$

$$\partial^\alpha K_F^0(\mathbf{x}) = \sum_{I \in F} 2^{-\epsilon(i_2 - i_1)}\, 2^{-\sum_{j=1}^n i_j(Q_j + [\![\alpha_j]\!])}\, \partial^\alpha [\varphi_0^I](2^{-I} \cdot \mathbf{x}).$$

It follows from Proposition 10.1 in Appendix II that (at least if $d_\ell < Q_1 + [\![\alpha_1]\!]$)

$$\big| \partial^\alpha K_F^\ell(\mathbf{x}) \big| \leq C\, N_1(\mathbf{x}_1)^{d_\ell} \prod_{j=1}^n [N_1(\mathbf{x}_1) + \cdots + N_j(\mathbf{x}_j)]^{-(Q_j + [\![\alpha_j]\!])}, \text{ and}$$

$$\big| \partial^\alpha K_F^0(\mathbf{x}) \big| \leq C\, N_1(\mathbf{x}_1)^\epsilon (N_1(\mathbf{x}_1) + N_2(\mathbf{x}_2))^{-\epsilon} \prod_{j=1}^n [N_1(\mathbf{x}_1) + \cdots + N_j(\mathbf{x}_j)]^{-(Q_j + [\![\alpha_j]\!])}.$$

The functions on the right hand side of the last two inequalities are integrable on $\mathbb{R}^N$. The proof of (2) then follows from the dominated convergence theorem:

$$\lim_{F \nearrow \mathcal{E}_n} \big\langle K_F, \psi \big\rangle = -\sum_{\ell=1}^{a_1} \int_{\mathbb{R}^N} K^\ell(\mathbf{x}) \frac{\partial \psi}{\partial x_\ell}(\mathbf{x})\, d\mathbf{x} + \int_{\mathbb{R}^N} K^0(\mathbf{x}) \psi(\mathbf{x})\, d\mathbf{x},$$

where

$$K^\ell(\mathbf{x}) = \sum_{I \in \mathcal{E}_n} 2^{i_1 d_\ell} [\varphi_\ell^I]_I(\mathbf{x}), \quad 1 \leq \ell \leq a_1,$$

$$K^0(\mathbf{x}) = \sum_{I \in \mathcal{E}_n} 2^{-\epsilon(i_2 - i_1)} [\varphi_0^I]_I.$$

6.3. Rewriting sums of bump functions with weak cancellation

It follows from Theorem 6.8 that a sum of dilates of normalized bump functions with weak cancellation converges to a flag kernel, and it follows from Theorem 6.1 that a flag kernel can be written as a sum of dilates of normalized bump functions with strong cancellation plus a sum of flag kernels adapted to strictly coarser flags. It follows that a sum of dilates of functions with *weak* cancellation can be rewritten as sums of dilates of functions with *strong* cancellation relative to coarser flags. In this section we give a direct proof of this fact. The basic idea is to use telescoping series to replace a function with weak cancellation by a sum of functions with strong cancellation plus an error term which belongs to a flag which is coarser than the original flag. Aside from its intrinsic interest, we shall need this observation in the forthcoming paper [10].

Thus consider the standard flag $\mathcal{F}_\mathcal{A}$ on $\mathbb{R}^N$ of step n associated to the decomposition

$$(\mathcal{A}) \quad \mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}.$$

Any strictly coarser flag $\mathcal{F}_\mathcal{B} \succ \mathcal{F}_\mathcal{A}$ then arises from a decomposition

$$(\mathcal{B}) \quad \mathbb{R}^N = \mathbb{R}^{b_1} \oplus \cdots \oplus \mathbb{R}^{b_m}$$

where $m < n$ and each $\mathbb{R}^{b_j} = \mathbb{R}^{a_{r_j}} \oplus \cdots \oplus \mathbb{R}^{a_{s_j}}$ where $1 = r_1$, $n = s_m$, $r_j \leq s_j$ for $1 \leq j \leq m$, and $r_{j+1} = s_j + 1$ for $1 \leq j \leq m-1$. As usual we let $\mathcal{E}_n$ denote the set of n-tuples of integers $I = (i_1, \ldots, i_n)$ with $i_1 \leq \cdots \leq i_n$. For any strictly coarser flag $\mathcal{F}_\mathcal{B}$ of step $m < n$ as above, we let $\mathcal{E}_\mathcal{B}$ denote the set of m-tuples of integers $J = (j_1, \ldots, j_m)$ with $j_1 \leq \cdots \leq j_m$. Given the argument for part (2) of Theorem 6.8, we shall only concern ourselves with finite sums, and thus will not need to worry about convergence questions.

Proposition 6.12. *Let $\mathcal{F}_\mathcal{A}$ denote the standard flag associated to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$. Suppose that*

$$(6.14) \qquad K(\mathbf{x}) = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I(\mathbf{x})$$

is a finite sum, where each φ^I is a normalized bump function for the flag $\mathcal{F}_\mathcal{A}$, with weak cancellation relative to $I \in \mathcal{E}_n$ with parameter ϵ. Then we can write

$$(6.15) \qquad K(\mathbf{x}) = \sum_{\mathcal{B} \succeq \mathcal{A}} \sum_{J \in \mathcal{E}_\mathcal{B}} \left[\eta_\mathcal{B}^J \right]_J (\mathbf{x}),$$

where the outer sum is taken over decompositions equal to or coarser than $\mathcal{A}$, and each $\eta_\mathcal{B}^J$ is a normalized bump function which has strong cancellation relative to the flag $\mathcal{F}_\mathcal{B}$ associated to the decomposition $\mathcal{B}$.

Proof. We argue by induction on the number of steps n in the original flag $\mathcal{A}$. For $n = 1$ there is nothing to prove, because then there is no distinction between weak and strong cancellation.

Thus suppose the proposition has been established for all flags of step less than or equal to $n-1$, and consider the flag of step n corresponding to the decomposition $\mathcal{A}$. The inductive step itself requires an induction. Let the function K be given by (6.14). Since each φ^I has weak cancellation with respect to I with parameter ϵ, we can write

$$\varphi^I = \sum_{S \subset \{1,2,\ldots,n-1\}} 2^{-\epsilon \sum_{l \in B}(j_{l+1}-j_l)} \, \eta_S^I$$

where the sum is over all subsets S of $\{1,\ldots,n-1\}$ and each η_S^I is a normalized bump function which has integral zero in each multi-variable $\mathbf{x}_r$ with $r \notin S$. Thus we have

$$(6.16) \qquad K = \sum_{I \in \mathcal{E}_n} \sum_{S \subset \{1,2,\ldots,n-1\}} 2^{-\epsilon \sum_{l \in S}(j_{l+1}-j_l)} \left[\eta_S^I\right]_I.$$

We will prove by induction on k, for $1 \le k \le n$, that K can be written

$$(6.17) \qquad K = \sum_{I \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k\}} 2^{-\varepsilon \sum_{l \in S}(j_{l+1}-j_l)} \left[\eta_S^I\right]_I + \sum_{\mathcal{B} \succ \mathcal{A}} K^{\mathcal{B}},$$

where the functions $\{\eta_S^I\}$ and $\{K^{\mathcal{B}}\}$ have the following properties:

(i) Each normalized bump function η_S^I has integral zero with respect to each variable $\mathbf{x}_r$ with $r \notin S$.

(ii) For each $\mathcal{B} \succ \mathcal{A}$, (*i.e.*, for each decomposition $\mathcal{B}$ strictly coarser than $\mathcal{A}$ and hence whose corresponding flag has step strictly less than n), the function $K^{\mathcal{B}}$ can be written as a finite sum

$$K^{\mathcal{B}} = \sum_{J \in \mathcal{E}(\mathcal{B})} [\theta_{\mathcal{B}}^J]_J,$$

where each $\theta_{\mathcal{B}}^J$ is a normalized bump function with weak cancellation with some parameter $\epsilon' > 0$ relative to the flag arising from the decomposition $\mathcal{B}$. (It follows from the induction hypothesis that each such function can be rewritten as a sum of dilates of normalized bump functions with strong cancellation.)

(iii) The bump functions $\{\eta_S^I\}$ and the $\{\theta_{\mathcal{B}}^J\}$ are uniformly normalized relative to the normalized bump functions $\{\varphi^I\}$ defined in (6.14).

Clearly equation (6.16) gives the desired conclusion in (6.17) for $k = 1$. Moreover, when $k = n$ the set S must be empty and thus we will have written K as an appropriate sum of dilates of normalized bump functions with strong cancellation.

Thus we turn to the induction step. Suppose that equation (6.17) holds for a given $k < n$. We must show that (6.17) also holds with k replaced by $k+1$. We split the first sum into two parts depending on whether or not the subset S

contains the element $n - k$:

$$K = \sum_{\substack{J \in \mathcal{E}_n \\ S \subseteq \{1,2,\ldots,n-k\} \\ n-k \in S}} 2^{-\varepsilon \sum_{l \in S}(j_{l+1}-j_l)}[\eta_S^J]_J + \sum_{\substack{I \in \mathcal{E}_n \\ S \subseteq \{1,2,\ldots,n-(k+1)\}}} 2^{-\varepsilon \sum_{l \in S}(j_{l+1}-j_l)}[\eta_S^I]_I + \sum_{\mathcal{B} \succ \mathcal{A}} K^{\mathcal{B}}$$

$$(6.18) \quad = K_1 + K_2 + K_3.$$

Now $K_2 + K_3$ is already of the form in (6.17) with k replaced by $k + 1$, so we only need to deal with K_1.

Thus let $n - k \in S \subset \{1, 2, \ldots, n - k\}$, and consider the corresponding term $\eta = \eta_S^J$ in K_1. Let Q_{n-k} denote the homogeneous dimension of the space $\mathbb{R}^{a_{n-k}}$. Then the function

$$(\eta_S^J)'(\mathbf{x}_1, \ldots, \mathbf{x}_n) = 2^{Q_{n-k}} \eta_S^J(\mathbf{x}_1, \ldots, 2 \cdot \mathbf{x}_{n-k}, \ldots, \mathbf{x}_n) - \eta_S^J(\mathbf{x}_1, \ldots, \mathbf{x}_{n-k}, \ldots, \mathbf{x}_n)$$

has cancellation in the variables $\mathbf{x}_r$ for all $r \notin S' = S \setminus \{n - k\}$. Note that $S' \subset \{1, 2, \ldots, n - (k+1)\}$. Let $J = (j_1, \ldots, j_n)$. Using a telescoping series we have

$$\eta_S^J(\mathbf{x}) = \sum_{i=1}^{j_{n-k+1}-j_{n-k}} 2^{-iQ_{n-k}}(\eta_S^J)'(\mathbf{x}_1, \ldots, 2^{-i} \cdot \mathbf{x}_{n-k}, \ldots, \mathbf{x}_n)$$
$$+ 2^{-(j_{n-k+1}-j_{n-k})Q_{n-k}} \eta_S^J(\mathbf{x}_1, \ldots, 2^{-(j_{n-k+1}-j_{n-k})} \cdot \mathbf{x}_{n-k}, \ldots, \mathbf{x}_n),$$

and hence
(6.19)
$$[\eta_S^J]_J = \sum_{i=j_{n-k}+1}^{j_{n-k+1}} \left[(\eta_S^J)'\right]_{(j_1,\ldots,j_{n-k-1},i,j_{n-k+1},\ldots,j_s)} + \left[\eta_S^J\right]_{(j_1,\ldots,j_{n-k-1},j_{n-k+1},j_{n-k+1},\ldots,j_s)}.$$

We regard the last term as associated to the coarser flag of step $(n-1)$ associated to the decomposition

$$(\mathcal{B}) \quad \mathbb{R}^{a_1} \oplus \mathbb{R}^{a_{n-k-1}} \oplus \left[\mathbb{R}^{a_{n-k}} \oplus \mathbb{R}^{a_{n-k+1}}\right] \oplus \mathbb{R}^{a_{n-k+2}} \oplus \cdots \oplus \mathbb{R}^{a_n}$$

where $\mathbb{R}^{a_{n-k}} \oplus \mathbb{R}^{a_{n-k+1}}$ is now considered one factor. If we set

$$J_i^{(n-k)} = (j_1, \ldots, j_{n-k-1}, i, j_{n-k+1}, \ldots, j_n), \, \widehat{J^{n-k}} = (j_1, \ldots, j_{n-k-1}, j_{n-k+1}, \ldots, j_n),$$

then $J_i^{(n-k)} \in \mathcal{E}_n$ for $j_{n-k} + 1 \leq i \leq j_{n-k+1}$ and $\widehat{J^{n-k}} \in \mathcal{E}_\mathcal{B}$. Moreover, η_S^J has weak cancellation with respect to the flag $(\mathcal{B})$ with parameter ϵ. Formula (6.19) then becomes

$$(6.20) \qquad [\eta_S^J]_J = \sum_{i=j_{n-k}+1}^{j_{n-k+1}} \left[(\eta_S^J)'\right]_{J_i^{(n-k)}} + \left[\eta_S^J\right]_{\widehat{J^{n-k}}}.$$

Applying the identity (6.20), we obtain

$$
\begin{aligned}
K_1 &= \sum_{J \in \mathcal{E}_n} \sum_{\substack{S \subseteq \{1,2,\ldots,n-k\} \\ n-k \in S}} 2^{-\varepsilon \sum_{l \in S}(j_{l+1}-j_l)} \left[\eta_S^J\right]_J \\
&= \sum_{J \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-1\}} 2^{-\varepsilon \sum_{l \in S \cup \{n-k\}}(j_{l+1}-j_l)} \left[\eta_{S \cup \{n-k\}}^J\right]_J \\
&= \sum_{J \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-1\}} 2^{-\varepsilon \sum_{l \in S \cup \{n-k\}}(j_{l+1}-j_l)} \sum_{i=j_{n-k+1}}^{j_{n-k}-1} \left[(\eta_S^{J_i^{(n-k)}})'\right]_{J_i^{(n-k)}} \\
&\quad + \sum_{J \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-1\}} 2^{-\varepsilon \sum_{l \in S \cup \{n-k\}}(j_{l+1}-j_l)} \left[\eta_S^{\widehat{J^{n-k}}}\right]_{\widehat{J^{n-k}}} \\
&= \Sigma_1 + \Sigma_2.
\end{aligned}
$$

In Σ_1 we change the order of summation, grouping together all the terms for which $J_i^{(n-k)}$ is a given $J' = (j_1', \ldots, j_n') \in \mathcal{E}_n$. Clearly, this condition forces J to differ from J' only in its component j_{n-k}, and we have $j_{n-k-1}' \leq j_{n-k} < j_{n-k}'$. In order to express the factor $2^{-\varepsilon \sum_{l \in S \cup \{n-k\}}(j_{l+1}-j_l)}$ in terms of J', we split the summation over the subsets S into two parts; the first consists of subsets S not containing $n-k-1$ and the second consists of the subsets which do contain $n-k-1$. We then have

$$
\begin{aligned}
\Sigma_1 &= \sum_{J' \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-2\}} 2^{-\varepsilon \sum_{l \in S}(j_{l+1}'-j_l')} \left(\sum_{i=j_{n-k-1}'}^{j_{n-k}'-1} 2^{-\varepsilon(j_{n-k+1}'-i)} \left[(\eta_S^{J'})'\right]_{J'} \right) \\
&\quad + \sum_{J' \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-2\}} 2^{-\varepsilon \sum_{l \in S}(j_{l+1}'-j_l')} 2^{-\varepsilon(j_{n-k+1}'-j_{n-k-1}')} \left(\sum_{i=j_{n-k-1}'}^{j_{n-k}'-1} \left[(\eta_{S \cup \{n-k-1\}}^{J'})'\right]_{J'} \right) \\
&= \Sigma_{1,1} + \Sigma_{1,2}.
\end{aligned}
$$

Each function appearing in $\Sigma_{1,1}$ has integral zero in each variable $\mathbf{x}_r$ with $r \notin B \subset \{1,\ldots,n-k-2\}$, and each function appearing in $\Sigma_{1,2}$ has integral zero in each variable $\mathbf{x}_r$ with $r \notin B \cup \{n-k-1\} \subset \{1,\ldots,n-k-1\}$. All the η' in the above formula are normalized relative to the initial data. Hence, the term in $\Sigma_{1,1}$ indexed by (J', S) contains a function $\tilde{\eta}_S^{J'}$ normalized relative to the initial data, and multiplied by a factor

$$
2^{-\varepsilon \sum_{l \in S}(j_{l+1}'-j_l')} \sum_{j_{n-k-1}'}^{j_{n-k}'-1} 2^{-\varepsilon(j_{n-k+1}'-i)} \lesssim 2^{-\varepsilon \sum_{l \in S}(j_{l+1}'-j_l')} ,
$$

whereas the corresponding term in $\Sigma_{1,2}$ is a bump function $\tilde{\eta}_{S \cup \{n-k-1\}}^{J'}$ normalized relative to the initial data and multiplied by

$$
2^{-\varepsilon \sum_{l \in S}(j_{l+1}'-j_l')} 2^{-\varepsilon(j_{n-k}'-j_{n-k-1}')}(j_{n-k}' - j_{n-k-1}') \lesssim 2^{-\varepsilon' \sum_{l \in S \cup \{n-k-1\}}(j_{l+1}'-j_l')} ,
$$

with $\varepsilon' < \varepsilon$. Hence,

$$\Sigma_1 = \sum_{J' \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-1\}} 2^{-\varepsilon \sum_{l \in S} (j'_{l+1} - j'_l)} \big[\tilde{\eta}_S^{J'} \big]_{J'} \ ,$$

and Σ_1 together with K_2 gives the first sum in (6.17) with k replaced by $k+1$.

It remains to prove that Σ_2 can be absorbed in the remainder term (second sum) of (6.17). We group together the terms at the same scale and separate the sets S containing $n-k-1$ from the others. Indexing the elements J' of $\mathcal{E}_\mathcal{B}$ as $J' = (j'_1, \ldots, j'_{n-k-1}, j'_{n-k}, \ldots, j'_{n-1}) \in \mathcal{E}_{n-1}$, we have

$$\Sigma_2 = \sum_{J \in \mathcal{E}_n} \sum_{S \subseteq \{1,2,\ldots,n-k-1\}} 2^{-\varepsilon \sum_{l \in S \cup \{n-k\}} (j_{l+1} - j_l)} \big[\eta_S^{\widehat{J^{n-k}}} \big]_{\widehat{J^{n-k}}}$$

$$= \sum_{J' \in \mathcal{E}_{n-1}} \sum_{S \subseteq \{1,2,\ldots,n-k-2\}} \bigg(\sum_{J \in \mathcal{E}_n : \widehat{J^{n-k}} = J'} 2^{-\varepsilon \sum_{l \in S \cup \{n-k\}} (j_{l+1} - j_l)} \big[\eta_S^{\widehat{J^{n-k}}} \big]_{\widehat{J^{n-k}}} \bigg)$$

$$+ \sum_{J' \in \mathcal{E}_{n-1}} \sum_{S \subseteq \{1,2,\ldots,n-k-2\}} \bigg(\sum_{J \in \mathcal{E}_n : \widehat{J^{n-k}} = J'} 2^{-\varepsilon \sum_{l \in S \cup \{n-k-1, n-k\}} (j_{l+1} - j_l)} \big[\eta_S^{\widehat{J^{n-k}}} \big]_{\widehat{J^{n-k}}} \bigg)$$

$$= \Sigma_{2,1} + \Sigma_{2,2} \ .$$

As in the previous discussion, each term in parentheses is a function normalized relative to the initial data, multiplied by a factor controlled by

$$2^{-\varepsilon \sum_{l \in S} (j_{l+1} - j_l)} \sum_{i=j'_{n-k-1}}^{j'_{n-k}} 2^{-\varepsilon(j'_{n-k} - i)} \lesssim 2^{-\varepsilon \sum_{l \in S} (j'_{l+1} - j'_l)} \ ,$$

for the terms in $\Sigma_{2,1}$ and by

$$2^{-\varepsilon \sum_{l \in S} (j_{l+1} - j_l)} 2^{-\varepsilon(j'_{n-k} - j'_{n-k-1})} (j'_{n-k} - j'_{n-k-1}) \lesssim 2^{-\varepsilon' \sum_{l \in S \cup \{n-k-1\}} (j'_{l+1} - j'_l)} \ ,$$

with $\varepsilon' < \varepsilon$, for the terms in $\Sigma_{2,2}$. It follows that

$$\Sigma_3 = \sum_{J' \in \mathcal{E}_{n-1}} \sum_{S \subseteq \{1,2,\ldots,n-k-1\}} 2^{-\varepsilon \sum_{l \in S} (j'_{l+1} - j'_l)} \big[\tilde{\eta}_S^{J'} \big]_{J'} \ ,$$

and this concludes the proof.   □

6.4. Restricted cancellation conditions

Our next result shows that the cancellation conditions in the definition of a flag kernel can be relaxed. As usual, let $\mathcal{F}$ denote the standard flag

$$(0) \subset \mathbb{R}^{a_n} \subset \mathbb{R}^{a_{n-1}} \oplus \mathbb{R}^{a_n} \subset \cdots \subset \mathbb{R}^{a_2} \oplus \cdots \oplus \mathbb{R}^{a_n} \subset \mathbb{R}^N \ .$$

Theorem 6.13. *Let $\mathcal{K}$ be a distribution in $\mathcal{S}(\mathbb{R}^N)$ which satisfies the conditions of Definition 2.3 for the flag $\mathcal{F}$, except that in condition (b), the values of the parameters $R_1, \ldots, R_s$ are restricted to $R_1 \geq R_2 \geq \cdots \geq R_s$. Then $\mathcal{K}$ is a flag kernel.*

Before giving the formal proof, let us explain what needs to be done. We are asserting that if the cancellation conditions in part (b) of Definition 2.3 are satisfied when $R_1 \geq R_2 \geq \cdots \geq R_s$, then they are satisfied for *all* values of the scaling parameters. To do this we fix a non-empty subset $I \subseteq \{1, \ldots, n\}$, and a constant $\rho_i > 0$ for every $i \in I$. For each $i \in I$ we choose $\psi_i \in C_0^\infty(\mathbb{R}^{a_i})$ equal to 1 on the N_i-ball of radius 1 and supported on the N_i-ball of radius 2, and with bounds on the norms $\{\|\psi_i\|_{(m)}\}$. We set $\mathbf{x}_I = (\mathbf{x}_i)_{i \in I}$ and $\mathbf{x}_I' = (\mathbf{x}_i)_{i \notin I}$, and put

$$\Psi_\rho(\mathbf{x}_I) = \prod_{i \in I} \psi_i(\rho_i \cdot \mathbf{x}_i) .$$

Let

$$\mathcal{K}_{\Psi,\rho}^\# = \mathcal{K}_{\Psi,\rho}^\#(\mathbf{x}_I') = \langle \mathcal{K}, \Psi_\rho \rangle$$

denote the distribution in the variables $\mathbf{x}_I'$ such that

$$\langle \mathcal{K}_{\Psi,\rho}^\#, \varphi \rangle = \langle \mathcal{K}, \Psi_\rho \otimes \varphi \rangle$$

for every test function φ in the variables $\mathbf{x}_I'$. We must then prove the following:

Suppose that $\mathcal{K}$ satisfies the hypotheses of Theorem 6.13. Then

(a) *If $I = \{1, \ldots, n\}$, then $|\langle \mathcal{K}, \Psi_\rho \rangle| \leq C$ where C is independent of the choice of $\{\rho_i\}$ and $\{\psi_i\}$.*

(b) *If I is a proper, non-empty subset of $\{1, \ldots, n\}$, let $i_0 = \min\{j \,\big|\, j \notin I\}$. Then $\mathcal{K}_{\Psi,\rho}^\#$ coincides with a smooth function for $\mathbf{x}_{i_0} \neq 0$, and for every multi-index $\bar\alpha$,*

$$(6.21) \qquad \left|\partial_{\mathbf{x}_I'}^{\bar\alpha} \mathcal{K}_{\Psi,\rho}^\#(\mathbf{x}_I')\right| \leq C_\alpha \prod_{i \notin I} \left(\sum_{\substack{l \notin I \\ l \leq i}} N_l(\mathbf{x}_l) \right)^{-Q_i - [\![\bar\alpha_i]\!]}$$

where $\{C_\alpha\}$ are independent of the choice of $\{\rho_i\}$ and $\{\psi_i\}$.

In proving (a) or (b), we may assume that $\mathcal{K}$ has compact support, and look for non-restricted cancellation estimates that only depend on the constants in Definition 4.1, and not on the size of the support. For general $\mathcal{K}$, the conclusion will then follow by a limiting argument, based on the following construction. We fix a C_0^∞-function φ on the real line, equal to 1 on a neighborhood of the origin, and set $\Phi = \varphi \otimes \cdots \otimes \varphi \in C_0^\infty(\mathbb{R}^N)$, $\Phi_r = \Phi \circ \delta_{r^{-1}}$. Then $\mathcal{K}_r = \Phi_r \mathcal{K}$ has compact support, satisfies the hypotheses of Theorem 6.13 uniformly in r, and $\lim_{r \to \infty} \mathcal{K}_r = \mathcal{K}$ in the sense of distributions.

Proof. For $i \in I$, define R_i as

$$(6.22) \qquad R_i = \max_{\substack{l \in I \\ l \geq i}} \rho_\ell .$$

Then $R_1 \geq R_2 \geq \cdots$, and so by hypothesis $\mathcal{K}_{\Psi,R}^\#$ satisfies the estimates in (6.21). Thus it suffices to prove that the difference $\mathcal{K}_{\Psi,\rho}^\# - \mathcal{K}_{\Psi,R}^\#$ does as well.

Denote by I^0 (respectively I^+) the set of $i \in I$ such that $R_i = \rho_i$ (respectively $R_i > \rho_i$). Setting

$$\eta_i(\mathbf{x}_i) = \psi_i(\rho_i \cdot \mathbf{x}_i) - \psi_i(R_i \cdot \mathbf{x}_i),$$

we have

$$
\begin{aligned}
\Psi_\rho(\mathbf{x}) - \Psi_R(\mathbf{x}) &= \Big(\prod_{i \in I^0} \psi_i(R_i \cdot \mathbf{x}_i) \Big) \Big(\prod_{i \in I^+} \psi_i(\rho_i \cdot \mathbf{x}_i) - \prod_{i \in I^+} \psi_i(R_i \cdot \mathbf{x}_i) \Big) \\
&= \Big(\prod_{i \in I^0} \psi_i(R_i \cdot \mathbf{x}_i) \Big) \Big(\prod_{i \in I^+} \big(\psi_i(R_i \cdot \mathbf{x}_i) + \eta_i(\mathbf{x}_i) \big) - \prod_{i \in I^+} \psi_i(R_i \cdot \mathbf{x}_i) \Big) \\
&= \Big(\prod_{i \in I^0} \psi_i(R_i \cdot \mathbf{x}_i) \Big) \sum_{\emptyset \neq J \subseteq I^+} \Big(\prod_{i \in J} \eta_i(\mathbf{x}_i) \Big) \Big(\prod_{i \in I^+ \setminus J} \psi_i(R_i \cdot \mathbf{x}_i) \Big) \\
&= \sum_{\emptyset \neq J \subseteq I^+} \Big(\prod_{i \in J} \eta_i(\mathbf{x}_i) \Big) \Big(\prod_{i \in I \setminus J} \psi_i(R_i \cdot \mathbf{x}_i) \Big) .
\end{aligned}
$$

Fix $J \subseteq I^+$, $J \neq \emptyset$. By definition of I^0 and I^+, for each $i \in J$, $R_i = \rho_l$ for some $l \in I^0$, $l > i$. Set $\bar{i} = \min\{l \in I^0 : l > i \text{ and } R_i = \rho_l\}$ and $\bar{J} = \{\bar{i} : i \in J\}$. By the cancellation of $\mathcal{K}$ in the variables $\mathbf{x}_i$ for $i \in I \setminus (J \cup \bar{J})$,

$$
\mathcal{K}^\#_{\Psi,\rho}(\mathbf{x}'_I) - \mathcal{K}^\#_{\Psi,R}(\mathbf{x}'_I) = \sum_{\emptyset \neq J \subseteq I^+} \Big\langle \mathcal{K}^\#_J, \Big(\prod_{i \in J} \eta_i \Big) \Big(\prod_{i \in \bar{J}} (\psi_i \circ \delta_{R_i}) \Big) \Big\rangle_{J \cup \bar{J}},
$$

where each $\mathcal{K}^\#_J$ is a distribution in the variables $\mathbf{x}_i$ with $i \in {}^cI \cup J \cup \bar{J}$, satisfying condition (a) of Definition 2.3.

Notice that this pairing can be expressed as an integral because the right-hand side f in the pairing above is supported where $\mathcal{K}^\#_J$ is smooth. To see this, denote by i_0 the smallest element of ${}^cI \cup J \cup \bar{J}$. Then $\mathcal{K}^\#_J$ is smooth for $\mathbf{x}_{i_0} \neq 0$. On the other hand, if $\mathbf{x} = (\mathbf{x}_i)_{i \in {}^cI \cup J \cup \bar{J}} \in \operatorname{supp} f$, all coordinates $\mathbf{x}_i$ are bounded away from zero except for those in $\bar{J}$. But every element of $\bar{J}$ is strictly larger than some element of J, therefore $i_0 \notin \bar{J}$.

We estimate the $\bar{\alpha}$-derivative of each $\mathcal{K}^\#_J$ by

$$
\begin{aligned}
&|\partial^{\bar{\alpha}}_{\mathbf{x}'_I} \mathcal{K}^\#_J(\mathbf{x}'_I, \mathbf{x}_J, \mathbf{x}_{\bar{J}})| \\
(6.23) \quad &\leq C_\alpha \Big(\prod_{i \in J} N_i(\mathbf{x}_i)^{-Q_i} \Big) \Big(\prod_{l \in \bar{J}} \big(\sum_{i \in J_l} N_i(\mathbf{x}_i) \big)^{-Q_l} \Big) \prod_{i \notin I} \Big(\sum_{\substack{l \notin I \\ l \leq i}} N_l(\mathbf{x}_l) \Big)^{-Q_i - [\![\bar{\alpha}_i]\!]},
\end{aligned}
$$

where $J_l = \{i \in J : \bar{i} = l\} = \{i \in J : R_i = R_l\}$. Since the J_l form a partition of J, we have

$$
|\mathcal{K}^\#_{\Psi,\rho}(\mathbf{x}'_I) - \mathcal{K}^\#_{\Psi,R}(\mathbf{x}'_I)| \leq C \Big(\prod_{i \notin I} \big(\sum_{\substack{l \notin I \\ l \leq i}} N_l(\mathbf{x}_l) \big)^{-Q_i - [\![\bar{\alpha}_i]\!]} \Big) \sum_{\emptyset \neq J \subseteq I^+} \prod_{l \in \bar{J}} V_l .
$$

With m_l denoting the cardinality of J_l,

$$V_l = \int_{\{N_i(\mathbf{x}_i) > R_l^{-1}, \forall i \in J_l\}} \int_{N_l(\mathbf{x}_l) < R_l^{-1}} \Big(\prod_{i \in J_l} N_i(\mathbf{x}_i)^{-Q_i} \Big) \Big(\sum_{i \in J_l} N_i(\mathbf{x}_i) \Big)^{-Q_l} d\mathbf{x}_l \prod_{i \in J_l} d\mathbf{x}_i$$

$$= R_l^{-Q_l} \int_{\{N_i(\mathbf{x}_i) > R_l^{-1}, \forall i \in J_l\}} \Big(\prod_{i \in J_l} N_i(\mathbf{x}_i)^{-Q_i} \Big) \Big(\sum_{i \in J_l} N_i(\mathbf{x}_i) \Big)^{-Q_l} \prod_{i \in J_l} d\mathbf{x}_i$$

$$\le C R_l^{-Q_l} \prod_{i \in J_l} \int_{N_i(\mathbf{x}_i) > R_l^{-1}} N_i(\mathbf{x}_i)^{-Q_i - Q_l/m_l} d\mathbf{x}_i \le C \ .$$

This concludes the proof.   □

6.5. Invariance of flag kernels under changes of variables

We study the effect of a change of variables on the class of flag distributions. If $\mathcal{K} \in \mathcal{S}'(\mathbb{R}^N)$, then formally $\langle \mathcal{K}, \psi \rangle = \int_{\mathbb{R}^N} K(\mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x}$, where K is the 'kernel' associated to $\mathcal{K}$. Let $F : \mathbb{R}^N \to \mathbb{R}^N$ be a diffeomorphism with inverse G. We want to define a new distribution $\mathcal{K}^\#$ which is the composition of $\mathcal{K}$ with the change of variables F. Now formally

$$\langle \mathcal{K} \circ F, \psi \rangle = \int_{\mathbb{R}^N} K(F(\mathbf{x})) \psi(\mathbf{x}) \, d\mathbf{x} = \int_{\mathbb{R}^N} K(\mathbf{y}) \psi(G(\mathbf{y})) \det(JG)(\mathbf{y}) \, dy$$

where JG is the Jacobian matrix of G. Thus if $\psi^\#(\mathbf{y}) = \psi(G(\mathbf{y})) \det(JG)(\mathbf{y})$, and if the change of variables has the property that $\psi \in \mathcal{S}(\mathbb{R}^N)$ implies $\psi^\# \in \mathcal{S}(\mathbb{R}^N)$, we can define $\mathcal{K}^\# = \mathcal{K} \circ F$ by setting $\langle \mathcal{K}^\#, \psi \rangle = \langle \mathcal{K}, \psi^\# \rangle$ for all $\psi \in \mathcal{S}(\mathbb{R}^N)$.

We are primarily interested in changes of variables of the form

$$F(x_1, \ldots, x_N) = (x_1 + P_1(\mathbf{x}), \ldots, x_N + P_N(\mathbf{x}))$$

where $P_1, \ldots, P_N$ are polynomials of the form $P_k(x_1, \ldots, x_N) = \sum c_k^\alpha x_1^{\alpha_1} \cdots x_{k-1}^{\alpha_{k-1}}$ with coefficients $c_k^\alpha \in \mathbb{R}$. Thus P_k depends only on the variables $\{x_1, \ldots, x_{k-1}\}$, and this guarantees that F is a diffeomorphism with inverse G of the same form. In particular, $\det(JG)(\mathbf{x})$ is a polynomial. Thus if $\psi \in \mathcal{S}(\mathbb{R}^N)$, then $\det(JG) \, \psi \circ G \in \mathcal{S}(\mathbb{R}^N)$, so $\mathcal{K}^\#$ is well-defined. We want to show that if $\mathcal{K}$ is a flag distribution adapted to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$, then $\mathcal{K}^\#$ has the same property. In order to show this, we need to make additional assumptions on the coefficients $\{c_k^\alpha\}$.

Definition 6.14. A change of variables $\mathbf{y} = F(\mathbf{x})$ of the form $y_k = x_k + P_k(\mathbf{x})$ where $P_k(\mathbf{x}) = \sum_{\alpha \in \mathcal{B}_k} c_k^\alpha x_1^{\alpha_1} \cdots x_{k-1}^{\alpha_{k-1}}$ is *allowable* if

$$\mathcal{B}_k = \Big\{ (\alpha_1, \ldots, \alpha_{k-1}) \in \mathbb{N}^{k-1} \mid \sum_{j=1}^{k-1} \alpha_j d_j = d_k \Big\}.$$

Theorem 6.15. *Let* $\mathcal{K} \in \mathcal{S}'(\mathbb{R}^N)$ *be a flag kernel adapted to the standard flag* $\mathcal{F}$ *coming from the decomposition* $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$. *If* $\mathbf{y} = F(\mathbf{x})$ *is an allowable change of variables, then* $\mathcal{K}^\# = \mathcal{K} \circ F$ *is a flag distribution for the same decomposition.*

Proof. We can assume that $\mathcal{K} = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$ where each φ^I is a normalized bump function having strong cancellation. Let $I \in E_N$. We consider the dilate $[\varphi^I]_I$ composed with an allowable change of variables. We have

$$[\varphi^I]_I(F(\mathbf{x})) = [\varphi^I]_I\big(\ldots, x_l + P_l(x_1, \ldots, x_{l-1}), \ldots \big)$$
$$= 2^{-\sum_{k=1}^{N} d_k i_k} \varphi^I\big(\ldots, 2^{-d_l i_l}[x_l + P_l(x_1, \ldots, x_{l-1})], \ldots \big)$$

Put

$$\theta^I(\mathbf{x}) = \varphi^I\big(\ldots, x_l + 2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1}), \ldots \big)$$

so that $[\theta^I]_I(\mathbf{x}) = [\varphi^I]_I(F(\mathbf{x}))$. Put

$$P_l^I(\mathbf{x}) = 2^{-d_l i_l} P_l(2^I \cdot \mathbf{x}) = \sum_{\mathcal{B}_l} 2^{-d_l i_l + \sum_{k=1}^{l-1} d_j \alpha_j i_j} c_k^\alpha x_1^{\alpha_1} \cdots x_{k-1}^{\alpha_{k-1}}.$$

Then $\theta^I(\mathbf{x}) = \varphi^I\big(x_1 + P_1^I(\mathbf{x}), \ldots, x_N + P_N^I(\mathbf{x}) \big)$. Since the change of variables is allowable and $I \in \mathcal{E}_n$, we have

$$-d_l i_l + \sum_{k=1}^{l-1} d_k \alpha_k i_k = -\sum_{k=1}^{l-1} \alpha_k d_k(i_l - i_k) \leq 0.$$

It follows that each P_l^I is normalized relative to P_l, and this shows that $[\varphi^I]_I(F(\mathbf{x})) = [\theta^I]_I(\mathbf{x})$, where $\theta^I \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ is normalized relative to φ^I.

Next we study the cancellation properties of θ^I. If we can show that each θ^I has weak cancellation relative to the multi-index $I \in \mathcal{E}_n$, it follows from Theorem 6.8 that

$$\mathcal{K} \circ F = \sum_I [\varphi^I]_I \circ F = \sum_I [\theta^I]_I$$

is a flag kernel, which is what we want to show.

To do this we use Proposition 5.7. Let $\{1, \ldots, n\} = A \cup B$ with $A \cap B = \emptyset$. We study

$$\int_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \theta^I(\mathbf{x}_A, \mathbf{x}_B) \, d\mathbf{x}_B$$
$$= \int_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \varphi^I\big(\ldots, x_l + 2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1}), \ldots \big) \, d\mathbf{x}_B.$$

Since φ^I has strong cancellation, we can write it as a sum of terms of the form $\partial_{j_1} \cdots \partial_{j_n} \widetilde{\varphi}^I$, where each index $j_l \in J_l$. It suffices to consider the integrals

$$\int_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \partial_{j_1} \cdots \partial_{j_n} \widetilde{\varphi}^I\big(\ldots, x_l + 2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 2^{d_{j_r} i_{j_r}} x_{j_r}, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1}), \ldots \big) \, d\mathbf{x}_B.$$

Let $r \in B$. In this last integral, replace the term

$$2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 2^{d_{j_r} i_{j_r}} x_{j_r}, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1})$$

for $l > j_r$ by the term $2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 0, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1})$; *i.e.*, we set $x_{j_r} = 0$ everywhere in the integrand except where it appears by itself in the j_r^{th} entry of $\partial_{j_1} \cdots \partial_{j_n} \widetilde{\varphi}^I$. Now

$$\int\limits_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \partial_{j_1} \cdots \partial_{j_n} \widetilde{\varphi}^I \left(\ldots, x_l + 2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 0, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1}), \ldots \right) d\mathbf{x}_B = 0$$

since one of the variables we integrate is x_{j_r}, and we are integrating the derivative of a Schwartz function. Thus it suffices to estimate the integral of the difference:

$$\int\limits_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \Big[\partial_{j_1} \cdots \partial_{j_n} \widetilde{\varphi}^I \left(\ldots, x_l + 2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 2^{d_{j_r} i_{j_r}} x_{j_r}, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1}), \ldots \right)$$

$$- \partial_{j_1} \cdots \partial_{j_n} \widetilde{\varphi}^I \left(\ldots, x_l + 2^{-d_l i_l} P_l(2^{d_1 i_1} x_1, \ldots, 0, \ldots, 2^{d_{l-1} i_{l-1}} x_{l-1}), \ldots \right) \Big] d\mathbf{x}_B.$$

A typical term in the polynomial P_l has the form $c_\alpha x_1^{\alpha_1} \cdots x_{l-1}^{\alpha_{l-1}}$ where $\alpha_1 d_1 + \cdots + \alpha_{l-1} d_{l-1} = d_l$. Thus we get a 'gain' whose size can be estimated by a sum of terms of the form

$$|c_\alpha| \, 2^{-d_l i_l + \alpha_1 d_1 i_1 + \cdots + \alpha_{l-1} d_{l-1} i_{l-1}} |x_1|^{\alpha_1} \cdots |x_{l-1}|^{\alpha_{l-1}}$$

where $\alpha_{j_r} > 0$. However,

$$-d_l i_l + \sum_{t=1}^{l-1} \alpha_t d_t i_t = \Big[-d_1 + \sum_{t=1}^{l-1} \alpha_t d_t \Big] i_l + \sum_{t=1}^{l-1} \alpha_t d_t (i_t - i_l)$$

$$= \sum_{t=1}^{l-1} \alpha_t d_t (i_t - i_l) \leq -\alpha_r d_r (i_l - i_r) \leq -\epsilon(i_{r+1} - i_r).$$

Thus for every $r \in B$ we have shown that

$$\left| \int_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \theta^I(\mathbf{x}_A, \mathbf{x}_B) \, d\mathbf{x}_B \right| \lesssim 2^{-\epsilon(i_{r+1} - i_r)}.$$

Thus with a smaller ϵ we have

$$\left| \int_{\bigoplus_{k \in B} \mathbb{R}^{a_k}} \theta^I(\mathbf{x}_A, \mathbf{x}_B) \, d\mathbf{x}_B \right| \lesssim \prod_{r \in B} 2^{-\epsilon(i_{r+1} - i_r)},$$

and it follows from Proposition 5.7 that θ^I has weak cancellation. This completes the proof.   □

6.6. Homogeneous nilpotent Lie groups

We now begin the study of operators $f \to f * \mathcal{K}$ where $\mathcal{K}$ is a flag kernel, and the convolution is on a homogeneous nilpotent Lie group G with Lie algebra $\mathfrak{g}$. To say that a Lie group G is homogeneous means that there is a one-parameter group of automorphisms $\delta_r : G \to G$ for $r > 0$, with $\delta_1 = \text{Id}$. As a manifold, G is an N-dimension real vector space, and we assume that with an appropriate choice of coordinates, $G = \mathbb{R}^N$ and the automorphisms are given by $\delta_r[\mathbf{x}] = r \cdot \mathbf{x} = (r^{d_1} x_1, \ldots, r^{d_N} x_N)$ with $1 \le d_1 \le d_2 \le \cdots \le d_N$. We begin by summarizing the facts about group multiplication, invariant vector fields, and group convolution that we need in this context. Additional background information, and in particular the proofs of formulas (6.24) and (6.25) below, can be found in the Chapter 1 of [3].

The product on $G = \mathbb{R}^N$ is given by a polynomial mapping; if $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$, the kth component of the product $\mathbf{xy}$ is given by

$$(6.24) \quad (\mathbf{xy})_k = x_k + y_k + M_k(\mathbf{x}, \mathbf{y}) = x_k + y_k + \sum_{\alpha, \beta \in \mathcal{M}_k} c_k^{\alpha, \beta} x_1^{\alpha_1} \cdots x_{k-1}^{\alpha_{k-1}} y_1^{\beta_1} \cdots y_{k-1}^{\beta_{k-1}}$$

where $\{c_k^{\alpha, \beta}\}$ are real constants, and

$$\mathcal{M}_k = \left\{ (\alpha; \beta) = (\alpha_1, \ldots, \alpha_{k-1}; \beta_1, \ldots, \beta_{k-1}) \,\Big|\, \sum_{l=1}^{k-1} d_l(\alpha_l + \beta_l) = d_k \right\}.$$

Note that

$$M_k(r \cdot \mathbf{x}, r \cdot \mathbf{y}) = r^{d_k} M_k(\mathbf{x}, \mathbf{y}).$$

Next, let $\{X_1, \ldots, X_N\}$ and $\{Y_1, \ldots, Y_N\}$ be the left- and right-invariant vector fields on G such that, at the origin, $X_k = Y_k = \partial_{x_k}$. Then

$$(6.25) \quad \begin{aligned} X_k &= \frac{\partial}{\partial x_k} + \sum_{\substack{l=k+1 \\ d_l > d_k}}^{N} P_{k_l}(\mathbf{x}) \frac{\partial}{\partial x_l} = \frac{\partial}{\partial x_k} + \sum_{\substack{l=k+1 \\ d_l > d_k}}^{N} \sum_{\alpha \in \mathfrak{H}_{d_l - d_k}} a_{k_l}^{\alpha} x_1^{\alpha_1} \cdots x_{l-1}^{\alpha_{l-1}} \frac{\partial}{\partial x_l}, \\ Y_k &= \frac{\partial}{\partial x_k} + \sum_{\substack{l=k+1 \\ d_l > d_k}}^{N} \widetilde{P}_{k_l}(\mathbf{x}) \frac{\partial}{\partial x_l} = \frac{\partial}{\partial x_k} + \sum_{\substack{l=k+1 \\ d_l > d_k}}^{N} \sum_{\alpha \in \mathfrak{H}_{d_l - d_k}} \widetilde{a}_{k_l}^{\alpha} x_1^{\alpha_1} \cdots x_{l-1}^{\alpha_{l-1}} \frac{\partial}{\partial x_l}, \end{aligned}$$

where $\{a_{k_l}^{\alpha}\}$ and $\{\widetilde{a}_{k_l}^{\alpha}\}$ are real constants, and the index set $\mathfrak{H}_d$ is defined in Proposition 2.1. It follows that $P_{k_l}, \widetilde{P}_{k_l} \in \mathcal{H}_{d_l - d_k}$.

The bi-invariant Haar measure on G is Lebesgue measure $d\mathbf{y} = dy_1 \cdots dy_N$. The *convolution* of functions $f, g \in L^1(G)$ is given by

$$f * g(\mathbf{x}) = \int_G f(\mathbf{xy}^{-1}) g(\mathbf{y}) \, d\mathbf{y} = \int_G f(\mathbf{y}) g(\mathbf{y}^{-1}\mathbf{x}) \, d\mathbf{y},$$

and the integral converges absolutely for almost all $x \in G$. The following result can be found on page 22 of [3].

Proposition 6.16. *Let $f, g \in \mathcal{C}^1(G) \cap L^1(G)$.*

(1) *If X is a left-invariant vector field and Y is a right invariant vector field, then*
    $X[f * g] = f * X[g]$ *and* $Y[f * g] = Y[f] * g$.

(2) *If X is a left-invariant vector field and $Y = \widetilde{X}$ is the unique right-invariant vector field agreeing with X at the origin, then $X[f] * g = f * Y[g]$.*

(3) *If $\delta = \delta_0$ denotes the Dirac delta-function at the origin, then $\varphi(x) = \varphi * \delta(x) = \delta * \varphi(x)$ for $\varphi \in \mathcal{C}_0^\infty(G)$. In particular, if X is a left-invariant vector field and Y is a right invariant vector field, $X[\varphi] = \varphi * X[\delta]$ and $Y[\varphi] = Y[\delta] * \varphi$.*

Using the formulas in (6.24), we can write the convolution of integrable functions f and g as

$$(6.26) \quad f * g(\mathbf{x}) = \int_{\mathbb{R}^N} f(\ldots, x_m - y_m - P_m(\mathbf{x}, \mathbf{y}), \ldots) \, g(\ldots, y_m, \ldots) \, dy_1 \cdots dy_N$$

where each P_m is a polynomial in the $2m-2$ variables $\{x_1, \ldots, x_{m-1}, y_1, \ldots, y_{m-1}\}$ satisfying $P_m(2^K \cdot \mathbf{x}, 2^K \cdot \mathbf{y}) = 2^{k_m d_m} P_m(\mathbf{x}, \mathbf{y})$. In the formula (6.26), the variables in $\mathbf{x}$ appear in the argument of f. However by a change of variables we can move some or all of them to the argument of g. Thus if S is any subset of $\{1, \ldots, N\}$, we can write

$$(6.27) \quad f * g(\mathbf{x}) = \int_{\mathbb{R}^N} f\big(u_1(\mathbf{x}, \mathbf{y}), \ldots, u_N(\mathbf{x}, \mathbf{y})\big) \, g\big(v_1(\mathbf{x}, \mathbf{y}), \ldots, v_N(\mathbf{x}, \mathbf{y})\big) \, dy_1 \cdots dy_N$$

where

$$(6.28) \quad \begin{aligned} u_m(\mathbf{x}, \mathbf{y}) &= \begin{cases} x_m - y_m - Q_m(\mathbf{x}, \mathbf{y}) & \text{if } m \in S, \\ y_m & \text{if } m \notin S, \end{cases} \\ v_m(\mathbf{x}, \mathbf{y}) &= \begin{cases} y_m & \text{if } m \in S, \\ x_m - y_m - Q_m(\mathbf{x}, \mathbf{y}) & \text{if } m \notin S. \end{cases} \end{aligned}$$

Here each $Q_m = Q_m^S$ is a polynomial in the variables $\{x_1, \ldots, x_{m-1}, y_1, \ldots, y_{m-1}\}$ with the same homogeneity as P_m; that is $Q_m(2^K \cdot \mathbf{x}, 2^K \cdot \mathbf{y}) = 2^{k_m d_m} Q_m(\mathbf{x}, \mathbf{y})$.

6.7. Support properties of convolutions $[\varphi]_I * [\psi]_J$

In this section we study the support properties of the convolution of dilates of normalized bump functions with compact support. Given integers $i, j \in \mathbb{Z}$, we set $i \vee j = \max\{i, j\}$. Given N-tuples $I = (i_1, \ldots, i_N)$, $J = (j_1, \ldots, j_N) \in \mathbb{Z}^N$, we set

$$(6.29) \qquad\qquad I \vee J = (i_1 \vee j_1, \ldots, i_N \vee j_N).$$

We want to show that the convolution $[\varphi]_I * [\psi]_J$ is the $I \vee J$-dilate of a normalized function.

Lemma 6.17. *Let $\varphi, \psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ have support in the ball $B(\rho)$. Then for any $I, J \in E_N$ there exists $\theta \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ supported in the ball $B(C\rho)$ such that $[\varphi]_I * [\psi]_J = [\theta]_{I \vee J}$, and $||\theta||_{(m)} \leq C_m ||\varphi||_{(m)} ||\psi||_{(m)}$. The constants C and $\{C_M\}$ can depend on the radius ρ, but are independent of the functions φ and ψ.*

Proof. Let $K = I \vee J$ and put

$$\theta = \big[[\varphi]_I * [\psi]_J\big]_{-K}.$$

It suffices to show that θ is supported in the set $\{\mathbf{x} \in \mathbb{R}^N : |x_k| \leq C\rho,\ 1 \leq k \leq N\}$ and that $||\theta||_{(m)} \leq C_m\, ||\varphi||_{(m)}||\psi||_{(m)}$ for some absolute constants C and $\{C_m\}$.

Making the change of variables $y_m \to 2^{k_m d_m} y_m$ and using the homogeneity of the functions $\{u_m\}$ and $\{v_m\}$, we have

$$
\begin{aligned}
\big[[\varphi]_I &* [\psi_J]\big]_{-K}(\mathbf{x}) \\
&= 2^{\sum_m d_m k_m}([\varphi]_I * [\psi]_J)(2^{d_1 k_1} x_1, \dots, 2^{d_N k_N} x_N) \\
&= 2^{\sum_m 2 d_m k_m} \int_{\mathbb{R}^N} [\varphi]_I\big(\dots, u_m(2^{d_1 k_1} x_1, \dots, 2^{d_N k_N} x_N, \mathbf{y}), \dots\big) \\
&\qquad\qquad \cdot [\psi]_J\big(\dots, v_m(2^{d_1 k_1} x_1, \dots, 2^{d_N k_N} x_N, \mathbf{y}), \dots\big) dy_1 \cdots dy_N \\
&= 2^{\sum_m d_m(2 k_m - i_m - j_m)} \int_{\mathbb{R}^N} \varphi\big(\dots, 2^{d_m(k_m - i_m)} u_m(\mathbf{x}, \mathbf{y}), \dots\big) \\
&\qquad\qquad \cdot \psi\big(\dots, 2^{d_m(k_m - j_m)} v_m(\mathbf{x}, \mathbf{y}), \dots\big) dy_1 \cdots dy_N.
\end{aligned}
$$

Note that $2^{d_m(k_m - i_m)} \geq 1$ and $2^{d_m(k_m - j_m)} \geq 1$. It follows that if $[\varphi_I * \psi_J]_{-K}(\mathbf{x}) \neq 0$, there exists $\mathbf{y} = (y_1, \dots, y_N) \in \mathbb{R}^N$ so that for $1 \leq m \leq N$,

$$
\begin{aligned}
(6.30) \qquad |u_m(\mathbf{x}, \mathbf{y})| &\leq 2^{d_m(k_m - i_m)} |u_m(\mathbf{x}, \mathbf{y})| \leq \rho, \\
|v_m(\mathbf{x}, \mathbf{y})| &\leq 2^{d_m(k_m - j_m)} |v_m(\mathbf{x}, \mathbf{y})| \leq \rho.
\end{aligned}
$$

We show by induction on m that these inequalities imply that $|x_m| + |y_m| \leq A_m \rho$ for an appropriate choice of constants $A_1 < A_2 < \cdots < A_N$. When $m = 1$, we have $|x_1 - y_1| \leq \rho$ and $|y_1| \leq \rho$, so $|x_1| + |y_1| \leq 3\rho$, and we can take $A_1 = 3$. Next, assume by induction that $|x_s| + |y_s| \leq A_s \rho$ for $1 \leq s < m$. Since $Q_m(\mathbf{x}, \mathbf{y})$ depends only on the variables $\{x_1, \dots, x_{m-1}, y_1, \dots, y_{m-1}\}$, it follows that $|Q_m(\mathbf{x}, \mathbf{y})| \leq B_m \rho$, where B_m is a constant that depends on the coefficients of the polynomial Q_m, on the constants $\{A_s\}$ for $s < m$, and on ρ. We have $|y_m| \leq \rho$ and $|x_m - y_m - Q_m(\mathbf{x}, \mathbf{y})| \leq \rho$, so $|x_m| + |y_m| \leq (B_m + 2)\rho$. This completes the proof of the statement about the support of θ.

To establish the estimate $||\theta||_{(m)} \leq C_m ||\varphi||_{(m)} ||\psi||_{(m)}$, we again use (6.27), but this time with the set $S = \{m \in \{1, \dots, N\} \,\big|\, j_m \leq i_m = k_m\}$. Of the two factors $\big\{2^{d_m(k_m - i_m)}, 2^{d_m(k_m - j_m)}\big\}$, the one which equals 1 multiplies the expression $x_m - y_m - Q_m(\mathbf{x}, \mathbf{y})$, while the term y_m is multiplied by the larger factor $2^{d_m(k_m - (i_m \wedge j_m))}$. Thus with this representation of the convolution $\varphi_I * \psi_J$, the integration takes place

over the set $E = \{\mathbf{y} \in \mathbb{R}^N : |y_m| \leq 2^{-d_m(k_m - (i_m \wedge j_m))}\}$. Thus we can estimate the size of $\left[\!\left[\varphi\right]_I * [\psi]_J\right]_{-K}(\mathbf{x})$ by

$$\left|\left[\varphi_I * \psi_J\right]_{-K}(\mathbf{x})\right| \leq 2^{\sum_m d_m(2k_m - i_m - j_m)} ||\varphi||_0 \, ||\psi||_0 \int_E d\mathbf{y} \leq C_0 \, ||\varphi||_0 \, ||\psi||_0$$

since $2k_m - i_m - j_m - k_m + (i_m \wedge j_m) = 0$. When we take derivatives of $\left[\varphi\right]_I * [\psi]_J\right]_{-K}(\mathbf{x})$, the terms involving the variables $\mathbf{x}$ are multiplied by the factor 1, and so we obtain in the same way the estimate

$$\left|\left|\left[\varphi_I * \psi_J\right]_{-K}\right|\right|_{(m)} \leq C_m \, ||\varphi||_{(m)} \, ||\psi||_{(m)}.$$

This completes the proof. $\qquad\square$

6.8. Decay and cancellation properties of convolutions $[\varphi]_I * [\psi]_J$

We want to study the decay and cancellation properties of the convolution $[\varphi]_I * [\psi]_J$ under the assumption that φ has cancellation in the variables $\{x_{l_1}, \ldots, x_{l_a}\}$ and ψ has cancellation in the variables $\{x_{m_1}, \ldots, x_{m_b}\}$. Here *decay* means that the size of $[\varphi]_I * [\psi]_J$ is small due to the difference between the N-tuples I and J; *cancellation* means that the integral of $[\varphi]_I * [\psi]_J$ with respect to some variables is zero. (See Section 5.2 and Definition 5.3 for the precise definition of strong and weak cancellation). Before stating our results, let us see what we should expect by considering the much simpler case in which the convolution $[\varphi]_I *_e [\psi]_J$ is taken with respect to the Abelian (Euclidean) vector space structure of $\mathbb{R}^N$ rather than the general homogeneous nilpotent Lie group structure G.

Let $I = (i_1, \ldots, i_N)$ and $J = (j_1, \ldots, j_N)$, and put

$$\begin{aligned} A_0 &= \{s \in \{1, \ldots, a\} \,|\, i_{l_s} \leq j_{l_s}\}, \\ A_1 &= \{s \in \{1, \ldots, a\} \,|\, i_{l_s} > j_{l_s}\} = \{1, \ldots, a\} \setminus A_0, \\ B_0 &= \{t \in \{1, \ldots, b\} \,|\, j_{m_t} \leq i_{m_t}\}, \\ B_1 &= \{t \in \{1, \ldots, b\} \,|\, j_{m_t} > i_{m_t}\} = \{m_1, \ldots, m_b\} \setminus B_0. \end{aligned}$$

Because of the hypothesis on cancellation, we can write

$$\begin{aligned} \varphi &= \partial_{l_1} \cdots \partial_{l_a} \widetilde{\varphi}, \\ \psi &= \partial_{m_1} \cdots \partial_{m_b} \widetilde{\psi}. \end{aligned}$$

For each $s \in A_0$ we can integrate by parts in the variable x_{l_s} in the integral $[\varphi]_I *_e [\psi]_J$, moving the derivative ∂_{l_s} from $\widetilde{\varphi}$ to $\widetilde{\psi}$. Since the width of the dilate $[\varphi]_I$ is narrower in this variable than the dilate $[\psi]_J$, this integration by parts gives a gain of $2^{-\epsilon(j_{l_s} - i_{l_s})}$, and we get such a gain for each $s \in A_0$. A similar argument shows that we get a gain of $2^{-\epsilon(i_{m_t} - j_{m_t})}$ for each $t \in B_0$. Thus the total gain from integration by parts is $\prod_{s \in A_0} 2^{-\epsilon(j_{l_s} - i_{l_s})} \prod_{t \in B_0} 2^{-\epsilon(i_{m_t} - j_{m_t})}$. In addition to this gain, we observe that in the convolution $[\varphi]_I *_e [\psi]_J$, the derivatives ∂_{l_s} for $s \in A_1$

and ∂_{m_t} for $t \in B_1$ can be pulled outside the integral. The final result is that there is a compactly supported function θ, normalized relative to φ and ψ, so that

$$[\varphi]_I *_e [\psi]_J = \prod_{s \in A_0} 2^{-\epsilon(j_{l_s} - i_{l_s})} \prod_{t \in B_0} 2^{-\epsilon(i_{m_t} - j_{m_t})} \prod_{s' \in A_1} \prod_{t' \in B_1} \partial_{l_{s'}} \partial_{m_{t'}} [\theta]_{I \vee J}.$$

In other words, we get exponential gains from variables where there is cancellation for the function with 'narrower' dilation, and the resulting convolution still has cancellation in the remaining variables.

When dealing with convolution on a homogeneous nilpotent Lie group, we cannot move Euclidean derivatives from one factor to the other. However, we can write Euclidean derivatives in terms of left- or right-invariant vector fields which can be moved across the convolution. But this process introduces error terms involving derivatives with respect to 'higher' variables, and these come with a gain involving the differences between entries of I or J. Thus in the case of nilpotent Lie groups, we might hope that convolution results in three kinds of terms: gains of the form $2^{-\epsilon|i_\ell - j_\ell|}$ coming from integration by parts in narrow variables with cancellation, residual cancellation of the convolution in some variables which are not used in the integration by parts, and finally gains of the type $2^{-\epsilon(i_{\ell+1} - i_\ell)}$ and $2^{-\epsilon(j_{\ell+1} - j_\ell)}$. This is in fact the case, and is made precise in the next lemma.

Suppose we are given two decompositions

$$(\mathcal{A}): \quad \mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n},$$
$$(\mathcal{B}): \quad \mathbb{R}^N = \mathbb{R}^{b_1} \oplus \cdots \oplus \mathbb{R}^{b_m}.$$

Let $\{J_1^{\mathcal{A}}, \ldots, J_n^{\mathcal{A}}\}$ be the indices corresponding to $(\mathcal{A})$ and let $\{J_1^{\mathcal{B}}, \ldots, J_m^{\mathcal{B}}\}$ be the indices corresponding to $(\mathcal{B})$. Define

$$\sigma : \{1, \ldots, N\} \to \{1, \ldots, n\} \quad \text{such that} \quad \ell \in J_{\sigma(\ell)}^{\mathcal{A}},$$
$$\tau : \{1, \ldots, N\} \to \{1, \ldots, m\} \quad \text{such that} \quad l \in J_{\tau(l)}^{\mathcal{B}}.$$

In what follows, $\pi_{\mathcal{A}}$ and $\pi_{\mathcal{B}}$ denote mappings from the set $\{1, \ldots, N\}$ to itself with the properties that $\pi_{\mathcal{A}}(\ell) \in J_{\sigma(\ell)}^{\mathcal{A}}$, and $\pi_{\mathcal{B}}(l) \in J_{\tau(l)}^{\mathcal{B}}$. Also recall from equation (4.5) in Section 4.1 that we can introduce mappings $p_{\mathcal{A}} : \mathcal{E}_n \to E_N$ and $p_{\mathcal{B}} : \mathcal{E}_m \to E_N$ so that

$$p_{\mathcal{A}}(i_1, \ldots, i_n) = (\overbrace{i_1, \ldots, i_1}^{a_1}, \overbrace{i_2, \ldots, i_2}^{a_2}, \ldots, \overbrace{i_n, \ldots, i_n}^{a_n}),$$
$$p_{\mathcal{B}}(j_1, \ldots, j_m) = (\overbrace{j_1, \ldots, j_1}^{b_1}, \overbrace{j_2, \ldots, j_2}^{b_2}, \ldots, \overbrace{j_m, \ldots, j_m}^{b_m}).$$

Lemma 6.18. *Suppose that $\varphi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ has cancellation in the variables x_ℓ for $\ell \in A \subset \{1, \ldots, N\}$, and that $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ has cancellation in the variables x_l for $l \in B \subset \{1, \ldots, N\}$. Let $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$ and $J = (j_1, \ldots, j_m) \in \mathcal{E}_m$. Set*

$$A_0 = \{\ell \in A \,|\, i_{\sigma(\ell)} \leq j_{\tau(\ell)}\},$$
$$B_0 = \{l \in B \,|\, j_{\tau(l)} \leq i_{\sigma(l)}\}.$$

Then $[\varphi]_I * [\psi]_J$ can be written as a sum of terms of the form

$$\prod_{\ell \in A_1} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in B_1} 2^{-\epsilon(i_{\tau(l)+1}-i_{\tau(l)})} \prod_{\ell \in A_2} 2^{-\epsilon(j_{\tau(\ell)}-i_{\sigma(\ell)})} \prod_{l \in B_2} 2^{-\epsilon(i_{\sigma(l)}-j_{\tau(l)})}$$

$$\left[\prod_{\ell \in A_3} \partial_{\pi_{\mathcal{A}}(\ell)} \prod_{l \in B_3} \partial_{\pi_{\mathcal{B}}(l)} \theta \right]_{p_{\mathcal{A}}(I) \vee p_{\mathcal{B}}(J)}$$

where θ is normalized relative to φ and ψ, $A = A_1 \cup A_2 \cup A_3$ and $B = B_1 \cup B_2 \cup B_3$ are disjoint unions, and we have $A_2 \subset A_0 \subset A_1 \cup A_2$ and $B_2 \subset B_0 \subset B_1 \cup B_2$. (We will have $\sigma(\ell) \neq n$ and $\tau(l) \neq n$.)

Proof. Using the cancellation hypotheses, it follows from Lemma 5.1 that we can write

$$\varphi = \left(\prod_{\ell \in A} \partial_{x_\ell} \right) [\varphi_A], \quad [\varphi]_I = \left(\prod_{\ell \in A} \left(2^{d_\ell i_{\sigma(\ell)}} \partial_{x_\ell} \right) \right) [\varphi_A]_I,$$

$$\psi = \left(\prod_{\ell \in B} \partial_{x_\ell} \right) [\psi_B], \quad [\psi]_J = \left(\prod_{\ell \in B} \left(2^{d_\ell j_{\tau(\ell)}} \partial_{x_\ell} \right) \right) [\psi_B]_J,$$

where φ_A is normalized relative to φ, and ψ_B is normalized relative to ψ. We can use Corollary 4.8 to write $[\varphi]_I$ as finite sums of terms of the form

$$\left(\prod_{\ell \in \widetilde{A}_1} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \right) \prod_{\ell \in A_1} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} Z_{\pi_{\mathcal{A}}(\ell)}) [\varphi_{A_1}]_I$$

where $A_1 \subset A$ is a possibly empty subset, $\widetilde{A}_1 = A \setminus A_1$, each $Z_{\pi_{\mathcal{A}}(\ell)}$ is either the corresponding left- or right-invariant vector field, and φ_{A_1} is normalized relative to φ. Moreover, according to the Remarks 4.9 following Corollary 4.8, the operators $\{Z_{\pi_{\mathcal{A}}(\ell)}\}$ can be put in any desired order. Similarly $[\psi]_J$ is a finite sum of terms of the form

$$\left(\prod_{l \in \widetilde{B}_1} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})} \right) \prod_{l \in B_1} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} Z_{\pi_{\mathcal{B}}(l)}) \left[\varphi_{B_1} \right]_J.$$

It follows that $[\varphi]_I * [\psi]_J$ is a finite sum of terms of the form

$$\prod_{\ell \in \widetilde{A}_1} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_1} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})}$$

$$\left(\prod_{\ell \in A_1} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} Z_{\pi_{\mathcal{A}}(\ell)}) [\varphi_{A_1}]_I * \prod_{l \in B_1} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} Z_{\pi_{\mathcal{B}}(l)}) [\varphi_{B_1}]_J \right).$$

Now let

$$A_2 = \{\ell \in A_1 \mid i_{\sigma(\ell)} \leq j_{\tau(\ell)}\} = A_1 \cap A_0, \quad \widetilde{A}_2 = \{\ell \in A_1 \mid i_{\sigma(\ell)} > j_{\tau(\ell)}\} = A_1 \setminus A_2,$$

$$B_2 = \{\ell \in B_1 \mid j_{\tau(l)} \leq i_{\sigma(l)}\} = B_1 \cap B_0, \quad \widetilde{B}'_2 = \{\ell \in B_1 \mid j_{\tau(l)} > i_{\sigma(l)}\} = B_1 \setminus B_2.$$

We choose Z_ℓ and Z_l as follows:

$$Z_\ell = \begin{cases} L_\ell & \text{if } \ell \in A_2, \\ R_\ell & \text{if } \ell \in \widetilde{A}_2, \end{cases} \qquad Z_l = \begin{cases} L_l & \text{if } l \in B_2, \\ R_l & \text{if } l \in \widetilde{B}_2. \end{cases}$$

Then since left-invariant vector fields commute with right invariant vector fields, we can use Proposition 6.16, part (1) and then Proposition 6.16, part (2) to write

$$\left(\prod_{\ell \in A_1} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} Z_{\pi_\mathcal{A}(\ell)})[\varphi_{A_1}]_I * \prod_{l \in B_1} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} Z_{\pi_\mathcal{B}(l)})[\varphi_{B_1}]_J \right)$$

$$= \prod_{\ell \in \widetilde{A}_2} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} R_{\pi_\mathcal{A}(\ell)}) \prod_{l \in \widetilde{B}_2} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} L_{\pi_\mathcal{B}(l)})$$

$$\left(\prod_{\ell \in A_2} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} L_{\pi_\mathcal{A}(\ell)})[\varphi_{A_1}]_I * \prod_{l \in B_2} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} R_l)[\psi_{B_1}]_J \right)$$

$$= \prod_{\ell \in \widetilde{A}_2} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} R_\ell) \prod_{l \in \widetilde{B}_2} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} L_{\pi_\mathcal{B}(l)})$$

$$\left([\varphi_{A_1}]_I * \prod_{\ell \in A_2} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} L_{\pi_\mathcal{A}(\ell)}) \prod_{l \in B_2} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} R_{\pi_\mathcal{B}(l)})[\psi_{B_1}]_J \right),$$

where φ_{A_1} and ψ_{B_1} are normalized[7] relative to φ and ψ. (Here $\prod_{\ell \in A_1'}$ is actually the product of the operators in the reverse order). Now we want to commute the operators

$$\prod_{\ell \in A_2} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} L_\ell) \quad \text{and} \quad \prod_{l \in B_2} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} R_l)$$

before applying them to $[\varphi_{B_1}]_J$. According to the third of the Remarks 4.9, the result is a sum of terms of the form

$$\prod_{\ell \in \widetilde{A}_3} 2^{-\epsilon(i_{\sigma(\ell)+1} - i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_3} 2^{-\epsilon(j_{\tau(l)+1} - j_{\tau(l)})} \prod_{l \in B_3} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} R_l) \prod_{\ell \in A_3} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} R_\ell)[\psi_{B_3}]$$

where

$$A_3 \subset A_2', \qquad \widetilde{A}_3 = A_2' \setminus A_3,$$
$$B_3 \subset B_2', \qquad \widetilde{B}_3 = B_2' \setminus B_3.$$

Thus

$$[\varphi_{A_1}]_I * \prod_{\ell \in A_2} (2^{d_{\pi_\mathcal{A}(\ell)} i_{\sigma(\ell)}} L_\ell) \prod_{l \in B_2} (2^{d_{\pi_\mathcal{B}(l)} j_{\tau(l)}} R_l)[\psi_{B_1}]_J$$

---

[7] The function φ_{A_1} also depends on A and ψ_{B_1} also depends on B. With minimal risk of confusion, we shall omit such notation.

is a sum of terms of the form

$$
\prod_{\ell \in \widetilde{A}_3} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_3} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})}
$$

$$
\left([\varphi_{A_1}]_I * \prod_{l \in B_3} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} R_l) \prod_{\ell \in A_3} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} R_\ell)[\psi_{B_3}] \right)
$$

$$
= \prod_{\ell \in \widetilde{A}_3} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_1''} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})}
$$

$$
\left(\prod_{l \in B_3} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} R_l)[\varphi_{A_1}]_I * \prod_{\ell \in A_3} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} R_\ell)[\psi_{B_3}] \right)
$$

$$
= \prod_{\ell \in \widetilde{A}_3} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_3} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})} \prod_{l \in B_3} 2^{-d_{\pi_{\mathcal{B}}(l)}(i_{\tau(l)}-j_{\tau(l)})}
$$

$$
\prod_{\ell \in A_3} 2^{-d_{\pi_{\mathcal{A}}(\ell)}(j_{\sigma(\ell)}-i_{\sigma(\ell)})} \left(\prod_{\widetilde{B}_3} (2^{d_{\pi_{\mathcal{B}}(l)} i_{\tau(l)}} R_l)[\varphi_{A_1}]_I * \prod_{\ell \in A_3} (2^{d_{\pi_{\mathcal{A}}(\ell)} j_{\sigma(\ell)}} R_\ell)[\psi_{B_3}] \right)
$$

$$
= \prod_{\ell \in \widetilde{A}_3} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_3} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})} \prod_{l \in B_3} 2^{-d_{\pi_{\mathcal{B}}(l)}(i_{\tau(l)}-j_{\tau(l)})}
$$

$$
\prod_{\ell \in A_3} 2^{-d_{\pi_{\mathcal{A}}(\ell)}(j_{\sigma(\ell)}-i_{\sigma(\ell)})} \left(\Big[\prod_{l \in B_3} R_l \varphi_{A_1} \Big]_I * \Big[\prod_{\ell \in A_3} R_\ell \psi_{B_3} \Big] \right).
$$

Now according to Lemma 6.17, we can write

$$
\Big[\prod_{\widetilde{B}_3} R_l \varphi_{A_1} \Big]_I * \Big[\prod_{\ell \in A_1''} R_\ell \psi_{B_3} \Big] = [\theta_{A_1,B_3}]_K
$$

where θ_{A_1,B_3} is normalized relative to φ and ψ and $K = I \vee J$. Thus it follows that $[\varphi]_I * [\psi]_J$ is a finite sum of terms of the form

$$
\prod_{\ell \in \widetilde{A}_1 \cup \widetilde{A}_3} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_1 \cup \widetilde{B}_3} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})} \prod_{\ell \in A_3} 2^{-d_{\pi_{\mathcal{A}}(\ell)}(j_{\sigma(\ell)}-i_{\sigma(\ell)})}
$$

$$
\prod_{l \in B_3} 2^{-d_{\pi_{\mathcal{B}}(l)}(i_{\tau(l)}-j_{\tau(l)})} \prod_{\ell \in \widetilde{A}_2} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} R_{\pi_{\mathcal{A}}(\ell)}) \prod_{l \in \widetilde{B}_2} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} L_{\pi_{\mathcal{B}}(l)})[\theta_{A_1,B_3}]_{I \vee J}.
$$

However, it follows from part (1) of Corollary 4.8 that we can write the product of vector fields $\prod_{\ell \in \widetilde{A}_2} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} R_\ell) \prod_{l \in \widetilde{B}_2} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} L_l)[\theta_{A_1,B_3}]_{I \vee J}$ as a sum of terms of the form

$$
\prod_{\ell \in \widetilde{A}_4} 2^{-\epsilon(i_{\sigma(\ell)+1}-i_{\sigma(\ell)})} \prod_{l \in \widetilde{B}_4} 2^{-\epsilon(j_{\tau(l)+1}-j_{\tau(l)})}
$$

$$
\prod_{\ell \in A_4} (2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} \partial_{\pi_{\mathcal{A}}}(\ell)) \prod_{l \in B_4} (2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} \partial_{\pi_{\mathcal{B}}(l)})[\theta_{A_4,B_4}]_{I \vee J}.
$$

Thus we have shown that $[\varphi]_I * [\psi]_J$ is a sum of terms of the form

$$\prod_{l \in \widetilde{B}_1 \cup \widetilde{B}_3 \cup B_4} 2^{-\epsilon(j_{\tau(l)+1} - j_{\tau(l)})} \prod_{\ell \in A_3} 2^{-d_{\pi_{\mathcal{A}}(\ell)}(j_{\sigma(\ell)} - i_{\sigma(\ell)})} \prod_{l \in B_3} 2^{-d_{\pi_{\mathcal{B}}(l)}(i_{\tau(l)} - j_{\tau(l)})}$$

$$\prod_{\ell \in A_4} \left(2^{d_{\pi_{\mathcal{A}}(\ell)} i_{\sigma(\ell)}} R_{\pi_{\mathcal{A}}(\ell)}\right) \prod_{l \in B_4} \left(2^{d_{\pi_{\mathcal{B}}(l)} j_{\tau(l)}} L_{\pi_{\mathcal{B}}(l)}\right) [\theta_{A_4, B_4}]_{I \vee J}.$$

This has the form asserted by the lemma, and so completes the proof. □

Corollary 6.19. *Let* $\varphi, \psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ *have strong cancellation relative to the same decomposition* $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$. *There exists* $\epsilon > 0$ *so that if* $I, J \in \mathcal{E}_n$, *it follows that* $[\varphi]_I * [\psi]_J$ *is a finite sum of terms of the form*

$$\prod_{\ell \in A} 2^{-\epsilon(i_{\ell+1} - i_\ell)} \prod_{\ell \in \widetilde{A}} 2^{-\epsilon|i_\ell - j_\ell|} \prod_{l \in B} 2^{-\epsilon(j_{l+1} - j_l)} \prod_{l \in \widetilde{B}} 2^{-\epsilon|i_l - j_l|} [\theta_{A,B}]_{I \vee J},$$

where,

(a) *the set* $\{1, \ldots, n\}$ *is the disjoint union of the sets* A *and* $\widetilde{A}$, *and of the sets* B *and* $\widetilde{B}$, *with* $n \notin A$ *and* $n \notin B$;

(b) *each function* $\theta_{A,B}$ *is normalized relative to* φ *and* ψ.

Proof. Let J_ℓ denote the set of subscripts ℓ such that $x_\ell \in \mathbb{R}^{a_\ell}$, and let

$$\sigma : \{1, \ldots, N\} \to \{1, \ldots, n\}$$

be the mapping such that $\sigma(\ell) \in J_\ell$ for all ℓ. Since φ and ψ both have strong cancellation relative to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$, if A_0 and B_0 are the sets defined in Lemma 6.18, it follows that $\{\sigma(\ell) \,|\, \ell \in A_0 \cup B_0\} = \{1, \ldots, n\}$. This means that the sets A_3 and B_3 of that Lemma must be empty, and the result follows. □

6.9. Truncated flag kernels

Definition 6.20. A flag distribution $\mathcal{K}$ adapted to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$ is a *truncated kernel of width* $a > 0$ if the differential inequalities given in part (a) of Definition 2.3 are replaced by

$$\left|\partial_{\mathbf{x}}^{\bar{\alpha}} K(\mathbf{x})\right| \le C_{\bar{\alpha}} \prod_{k=1}^n \left[a + N_1(\mathbf{x}_1) + \cdots + N_k(\mathbf{x}_k)\right]^{-Q_k - [\![\bar{\alpha}_k]\!]}.$$

$\mathcal{K}$ is an *improved truncated kernel* if it is a truncated kernel, and in addition satisfies

$$\left|\partial_{\mathbf{x}}^{\bar{\alpha}} K(\mathbf{x})\right| \le C_{\bar{\alpha}} \frac{a}{a + N_1(\mathbf{x}_1)} \prod_{k=1}^n \left[a + N_1(\mathbf{x}_1) + \cdots + N_k(\mathbf{x}_k)\right]^{-Q_k - [\![\bar{\alpha}_k]\!]}.$$

Our objective is to establish the following:

Proposition 6.21. *Let $\mathcal{K}$ be a flag distribution.*

(1) *If $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ has support in the unit ball, then $\mathcal{K} * \psi$ and $\psi * \mathcal{K}$ are truncated flag kernels of width 1.*

(2) *If $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ has support in the unit ball, and if $\int_{\mathbb{R}^N} \psi(\mathbf{x})\,d\mathbf{x} = 0$, then $\mathcal{K} * \psi$ and $\psi * \mathcal{K}$ are improved truncated flag kernels of width 1.*

Proof. We can write $\mathcal{K} = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I + \sum_{j=1}^r \mathcal{K}_j$, where $\{\varphi^I\}$ are normalized unit bump functions with strong cancellation, and $\{\mathcal{K}_1, \ldots, \mathcal{K}_r\}$ are flag distributions adapted to coarser flags. If $\mathcal{K}_0 = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$, it suffices to show that the proposition is true for $\mathcal{K}_0$. We consider $\psi * \mathcal{K}_0$. The case of $\mathcal{K}_0 * \psi$ is handled similarly. We have

$$\psi * \mathcal{K}_0 = \sum_{I \in \mathcal{E}_n} \psi * [\varphi^I]_I = \sum_{I \in \mathcal{E}_n^-} \psi * [\varphi^I]_I + \sum_{k=1}^{n-1} \sum_{I \in \mathcal{E}_n^k} \psi * [\varphi^I]_I + \sum_{I \in \mathcal{E}_n^+} \psi * [\varphi^I]_I$$

$$= \mathrm{I} + \sum_{k=1}^{n-1} \mathrm{II}_k + \mathrm{III}$$

where

$$\mathcal{E}_n^- = \{I = (i_1, \ldots, i_n) \in \mathcal{E}_n \,\big|\, i_n \le 0\},$$
$$\mathcal{E}_n^k = \{I = (i_1, \ldots, i_n) \in \mathcal{E}_n \,\big|\, i_k \le 0 < i_{k+1}\},$$
$$\mathcal{E}_n^+ = \{I = (i_1, \ldots, i_n) \in \mathcal{E}_n \,\big|\, 0 < i_1\}.$$

Denote the element $(0, \ldots, 0) \in \mathcal{E}_n$ by $\mathbf{0}$. If $I = (i_1, \ldots, i_n) \in \mathcal{E}_n^k$, then $i_1 \le i_2 \le \cdots \le i_k \le 0$, and we put $\widetilde{I}_k = (0, \ldots, 0, i_{k+1}, \ldots, i_n)$. Then

$$I \vee \mathbf{0} = \begin{cases} \mathbf{0} & \text{if } I \in \mathcal{E}_n^-, \\ \widetilde{I}_k & \text{if } I \in \mathcal{E}_n^k, \\ I & \text{if } I \in \mathcal{E}_n^+. \end{cases}$$

Let

$$\widetilde{\mathcal{E}}_n^k = \{I \in \mathcal{E}_n \,\big|\, i_1 = \cdots = i_k = 0\}.$$

Note that if $I \in \mathcal{E}_n^k$ then $\widetilde{I}_k \in \widetilde{\mathcal{E}}_n^k$. According to Lemmas 6.17 and 6.18, each term $\psi * [\varphi^I]_I$ has weak cancellation. Moreover, for each $I \in \mathcal{E}_n$ there exists $\theta^I \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, normalized with respect to ψ and φ^I, so that

$$(6.31) \quad \begin{aligned} I \in \mathcal{E}_n^- &\implies \psi * [\varphi^I]_I = 2^{-\epsilon(|i_1| + \cdots + |i_n|)} [\theta^I]_{\bar{0}}, \\ I \in \mathcal{E}_n^k &\implies \psi * [\varphi^I]_I = 2^{-\epsilon(|i_1| + \cdots + |i_k|)} [\theta^I]_{\widetilde{I}_k}, \\ I \in \mathcal{E}_n^+ &\implies \psi * [\varphi^I]_I = [\theta^I]_I. \end{aligned}$$

We have

$$\mathrm{I}(\mathbf{x}) = \sum_{I \in \mathcal{E}_n^-} 2^{-\epsilon(|i_1| + \cdots + |i_n|)} [\theta^I]_{\mathbf{0}}(\mathbf{x}) = \sum_{I \in \mathcal{E}_n^-} 2^{-\epsilon(|i_1| + \cdots + |i_n|)} \theta^I(\mathbf{x}) = \widetilde{\theta}^{\mathbf{0}}(\mathbf{x}),$$

where the series converges to a normalized unit bump function $\widetilde{\theta}^{\mathbf{0}}$. Next, we can write

$$\mathrm{II}_k(\mathbf{x}) = \sum_{I \in \mathcal{E}_n^k} \psi * [\varphi^I]_I(\mathbf{x}) = \sum_{\substack{J \in \widetilde{\mathcal{E}}_n^k}} \Big[\sum_{\substack{I \in \mathcal{E}_n^k \\ \widetilde{I}_k = J}} \psi * [\varphi^I]_I \Big](\mathbf{x})$$

$$= \sum_{\substack{J \in \widetilde{\mathcal{E}}_n^k}} \Big[\sum_{\substack{I \in \mathcal{E}_n^k \\ \widetilde{I}_k = J}} 2^{-\epsilon(|i_1| + \cdots + |i_k|)} [\theta^I]_{\widetilde{I}_k} \Big](\mathbf{x})$$

$$= \sum_{\substack{J \in \widetilde{\mathcal{E}}_n^k}} \Big[\sum_{\substack{I \in \mathcal{E}_n^k \\ \widetilde{I}_k = J}} 2^{-\epsilon(|i_1| + \cdots + |i_k|)} \theta^I \Big]_J(\mathbf{x}) = \sum_{\substack{J \in \widetilde{\mathcal{E}}_n^k}} \big[\widetilde{\theta}^J \big]_J(\mathbf{x}),$$

where

$$\sum_{\substack{I \in \mathcal{E}_n^k \\ \widetilde{I}_k = J}} 2^{-\epsilon(|i_1| + \cdots + |i_k|)} \theta^I = \widetilde{\theta}^J$$

converges to a normalized unit bump function. Thus each II_k is a flag kernel with a decomposition into dilates of normalized bump functions where all of the dilation parameters are non-negative. This is also true of the term III.

Thus we have shown that

$$\psi * \mathcal{K}_0 = \sum_{I \in \mathcal{E}_n^+} [\widetilde{\theta}^I]_I$$

where

$$\mathcal{E}_n^+ = \big\{ I = (i_1, \ldots, i_n) \in \mathbb{Z}^n \,\big|\, 0 \le i_1 \le i_2 \le \cdots \le i_n \big\}.$$

But then it follows from the second inequality in Proposition 10.1, as in the proof of Proposition 6.9 that $\psi * \mathcal{K}_0$ satisfies the differential inequalities of a truncated kernel of width one.

If we assume in addition that $\int_{\mathbb{R}^N} \psi(\mathbf{x}) \, d\mathbf{x} = 0$, then we can write $\psi = \sum_{l=1}^N \psi_l$, with

$$\int_{\mathbb{R}} \psi_l(x_1, \ldots, x_{l-1}, t, x_{l+1}, \ldots, x_N) \, dt = 0.$$

We can repeat the argument given above with ψ replaced by ψ_l. It then follows from Lemma 6.18 that instead of the formulas given in equation (6.31), we get

$$(6.32) \quad \begin{aligned} I \in \mathcal{E}_n^- &\implies \psi_l * [\varphi^I]_I = 2^{-\epsilon(|i_1| + \cdots + |i_n|)} [\theta^I]_{\bar{0}}, \\ I \in \mathcal{E}_n^k &\implies \psi_l * [\varphi^I]_I = 2^{-\epsilon(|i_1| + \cdots + |i_k|)} [\theta^I]_{\widetilde{I}_k} && \text{if } k < l, \\ I \in \mathcal{E}_n^k &\implies \psi_l * [\varphi^I]_I = 2^{-d_l i_l} 2^{-\epsilon(|i_1| + \cdots + |i_k|)} [\theta^I]_{\widetilde{I}_k} && \text{if } k \ge l, \\ I \in \mathcal{E}_n^+ &\implies \psi_l * [\varphi^I]_I = 2^{-d_l i_l} [\theta^I]_I. \end{aligned}$$

Again using the second inequality in Proposition 10.1, and observing that the case $l = 1$ gives the worst estimate, we see that $\psi_l * \mathcal{K}_0$ satisfies the differential inequalities of an improved truncated kernel of width one. $\qquad \square$

Remark 6.22. We point out that we can relax the $\mathcal{C}^\infty$ requirement on the function ψ in Proposition 6.21 in the following way. An examination of the arguments in Sections 6.6 and 6.8 and the proof just given shows that for any integer m there exists an integer M, so that $\mathcal{K} * \psi$ and $\psi * \mathcal{K}$ satisfy the required differential inequalities and cancellation properties for orders of differentiation not exceeding m, if ψ is supposed to be of class $\mathcal{C}^M$.

7. Convolution of flag kernels

Let $\mathcal{K} \in \mathcal{S}'(\mathbb{R}^N)$ be a flag distribution on the homogeneous nilpotent Lie group $G = \mathbb{R}^N$, adapted to the standard flag associated with the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$. Define a left-invariant operator $T_\mathcal{K} : \mathcal{S}(\mathbb{R}^N) \to \mathcal{C}^\infty(\mathbb{R}^N)$ by setting

$$T_\mathcal{K}[\phi](\mathbf{x}) = \phi * \mathcal{K}(\mathbf{x}) = \langle \mathcal{K}, \phi_\mathbf{x}^\# \rangle$$

where, if $\phi \in \mathcal{S}(\mathbb{R}^N)$, we set $\phi_\mathbf{x}^\#(\mathbf{y}) = \phi(\mathbf{x}\mathbf{y}^{-1})$. If $\mathcal{K}_1$ and $\mathcal{K}_2$ are two flag kernels on G, we want to make sense of the composition $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}$, and show that the resulting operator is of the form $T_{\mathcal{K}_3}$ where $\mathcal{K}_3$ is a third flag kernel on G. Now formally

$$T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}[\phi] = (T_{\mathcal{K}_1}[\phi]) * \mathcal{K}_2 = (\phi * \mathcal{K}_1) * \mathcal{K}_2 = \phi * (\mathcal{K}_1 * \mathcal{K}_2),$$

so the operator $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}$ should be given by convolution with the distribution $\mathcal{K}_1 * \mathcal{K}_2$. However we cannot directly define the composition $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}[\phi] = T_{\mathcal{K}_2}(T_{\mathcal{K}_1}[\phi])$, even if $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, since $T_{\mathcal{K}_1}[\phi]$ need not belong to $\mathcal{S}(\mathbb{R}^N)$. Also, in general one cannot convolve an arbitrary pair of distributions unless one of them has compact support.

We will define the convolution $\mathcal{K}_1 * \mathcal{K}_2$ somewhat indirectly. In Section 7.1 we show that if $\phi \in \mathcal{S}(\mathbb{R}^N)$, then $T_\mathcal{K}[\phi] \in L^2(\mathbb{R}^N)$ and the mapping $T_\mathcal{K} : \mathcal{S}(\mathbb{R}^N) \to L^2(\mathbb{R}^N)$ has a (unique) continuous extension to a mapping of $L^2(\mathbb{R}^N)$ to itself. This allows us to define $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}$ as the composition of two mappings from $L^2(\mathbb{R}^N)$ to itself. Then in Section 7.2, we show that this composition is given by convolution with a distribution which is given as a sum of convolutions of dilates of bump functions. The key is then to recognize this sum as a flag kernel. The combinatorics are rather complicated, so in Section 7.3 we work out an explicit example. In Section 7.4 we prove the main result, Theorem 7.4, which shows that the convolution of two flag kernels is a sum of flag kernels. Finally in Section 7.5 we work out some additional examples.

7.1. Boundedness on L^2

In this section we show that convolution with a flag kernel extends to a bounded operator on $L^2(\mathbb{R}^N)$. Later in Section 8 we will see more: such operators are bounded on $L^p(\mathbb{R}^N)$ for $1 < p < \infty$.

Lemma 7.1. *Let $\mathcal{K}$ be a flag kernel on $\mathbb{R}^N$. Then there is a constant C so that if $T_\mathcal{K}[\phi] = \phi * \mathcal{K}$ for $\phi \in \mathcal{S}(\mathbb{R}^N)$ then $||T_\mathcal{K}[\phi]||_{L^2(\mathbb{R}^N)} \leq C \, ||\phi||_{L^2(\mathbb{R}^N)}$. As a consequence, there is a unique extension of $T_\mathcal{K}$ to a bounded operator from $L^2(\mathbb{R}^N)$ to itself.*

Proof. Using Corollary 6.2, we can assume that $\mathcal{K}$ is a flag kernel adapted to a standard flag $\mathcal{F}$ of length n as given in equation (2.4), and that there is a uniformly bounded family of functions $\{\varphi^I\} \subset \mathcal{C}_0^\infty(\mathbb{R}^N)$, each having strong cancellation relative to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$ such that $\mathcal{K} = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$ with convergence in the sense of distributions.

For any $I \in \mathcal{E}_n$, let $T_I[f] = T_{\varphi^I}[f] = f * [\varphi^I]_I$. Then $||T_I[f]||_{L^2} \le ||\varphi^I||_{L^1} \, ||f||_{L^2}$. If $(f, g) = \int_{\mathbb{R}^N} f(\mathbf{x}) \overline{g(\mathbf{x})} \, d\mathbf{x}$ is the standard inner product in $L^2(\mathbb{R}^N)$, it follows from Fubini's theorem that $\big(T_I[f], g \big) = \big(f, \widetilde{T}_I[g] \big)$ where $T_{\widetilde{I}}[g] = g * [\widetilde{\varphi}^I]_I$ and $\widetilde{\varphi}^I(\mathbf{x}) = \overline{\varphi^I(\mathbf{x}^{-1})}$. Thus the Hilbert space adjoint of the operator T_{φ^I} is the operator $T_\varphi^* = T_{\widetilde{\varphi}^I}$. Fubini's theorem also shows that $T_I \circ T_J^* = T_{\widetilde{\varphi}^I * \varphi^J}$ and $T_J^* \circ T_I = T_{\varphi^I * \widetilde{\varphi}^J}$. Thus if $I, J \in \mathcal{E}_n$, the L^2-norm of the operators $T_I \circ T_J^*$ and $T_J^* \circ T_I$ are bounded by the L^1 norms of $[\widetilde{\varphi}^J]_J * [\varphi^I]_I$ and $[\varphi^I]_I * [\widetilde{\varphi}^J]_J$. It follows from Corollary 6.19 that

$$(7.1) \qquad \begin{aligned} ||[\widetilde{\varphi}]_J * [\varphi]_I||_{L^1(\mathbb{R}^N)} &+ ||[\varphi]_I * [\widetilde{\varphi}]_J||_{L^1(\mathbb{R}^N)} \\ &\le C \, 2^{-\epsilon|i_n - j_n|} \prod_{\ell=1}^{n-1} \big[2^{-\epsilon|i_\ell - j_\ell|} + \min\big\{ 2^{-\epsilon(i_{\ell+1} - i_\ell)}, 2^{-\epsilon(j_{\ell+1} - j_\ell)} \big\} \big]. \end{aligned}$$

For any finite subset $F \subset \mathcal{E}_n$, set $\mathcal{K}_F(\mathbf{x}) = \sum_{I \in F} [\varphi^I]_I(\mathbf{x})$. Then for any $\phi \in \mathcal{S}(\mathbb{R}^N)$, $\langle \mathcal{K}, \phi \rangle = \lim_{F \nearrow \mathcal{E}_n} \langle \mathcal{K}_F, \phi \rangle$, and in particular, if $\phi_\mathbf{x}^\#(\mathbf{y}) = \phi(\mathbf{x}\mathbf{y}^{-1})$,

$$T_{\mathcal{K}}[\phi](\mathbf{x}) = \big\langle \mathcal{K}, \phi_\mathbf{x}^\# \big\rangle = \lim_{F \nearrow \mathcal{E}_n} \big\langle \mathcal{K}_F, \phi_\mathbf{x}^\# \big\rangle = \lim_{F \nearrow \mathcal{E}_n} \sum_{I \in F} \phi * [\varphi^I]_I(\mathbf{x}) = \lim_{F \nearrow \mathcal{E}_n} \sum_{I \in F} T_I[\phi](\mathbf{x}).$$

It follows from the almost orthogonality estimate in (7.1) and the Cotlar–Stein Theorem (see for example [11], page 280) that there is a constant C independent of the finite set F such that

$$\Big|\Big| \sum_{I \in F} T_{[\varphi^I]_I}[\phi] \Big|\Big|_{L^2} \le C \, ||\phi||_{L^2}.$$

But then Fatou's lemma implies that $||T_{\mathcal{K}}[\phi]||_{L^2} \le C \, ||\phi||_{L^2}$ for all $\phi \in \mathcal{S}(\mathbb{R}^N)$. This completes the proof. $\qquad \square$

Corollary 7.2. *Supposet that $\mathcal{K}$ is a flag kernel, and that $\mathcal{K} = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$. Then for all $f \in L^2(\mathbb{R}^N)$,*

$$\lim_{F \nearrow \mathcal{E}_n} \Big|\Big| \sum_{I \in F} T_{[\varphi]_I}[f] - T_{\mathcal{K}}[f] \Big|\Big|_{L^2} = 0.$$

Proof. Since $\mathcal{S}(\mathbb{R}^N)$ is dense in $L^2(\mathbb{R}^N)$, and since $T_{\mathcal{K}}$ is bounded on $L^2(\mathbb{R}^N)$, it suffices to show that $\lim_{K \nearrow \mathcal{E}_n} ||\sum_{I \in F} T_{[\varphi^I]_I}[\psi] - T_{\mathcal{K}}[\psi]||_{L^2} = 0$ for $\psi \in \mathcal{S}(\mathbb{R}^N)$. (Both $T_{\mathcal{K}}$ and $\sum_{I \in F} T_{[\varphi^I]_I}$ are bounded on $L^2(\mathbb{R}^N)$ with norm independent of F). But for $\psi \in \mathcal{S}(\mathbb{R}^N)$, the result follows from Theorem 6.8 and the discussion following it on page 663. This completes the proof. $\qquad \square$

7.2. Composition of convolution operators

Let $\mathcal{K}_j$, $j = 1, 2$, be two flag kernels on $\mathbb{R}^N$, and let $T_{\mathcal{K}_j}[\phi] = \phi * \mathcal{K}_j$ be the corresponding convolution operators. According to Lemma 7.1, each of these operators is bounded on $L^2(\mathbb{R}^N)$, and hence the composition $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}$ is well-defined as a bounded operator on $L^2(\mathbb{R}^N)$. Our main result is the following.

Theorem 7.3. *Let $\mathcal{F}_1, \mathcal{F}_2$ be two standard flags on $\mathbb{R}^N$, and let $\mathcal{F}_0$ be the coarsest flag on $\mathbb{R}^N$ which is finer than both $\mathcal{F}_1$ and $\mathcal{F}_2$. For $j = 1, 2$, let $\mathcal{K}_j$ be a flag kernel adapted to the flag $\mathcal{F}_j$. Then $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}$ is a flag kernel adapted to the flag $\mathcal{F}_0$.*

In order to study the composition $T_{\mathcal{K}_2} \circ T_{\mathcal{K}_1}$, we want to relate it to the decompositions of $\mathcal{K}_1$ and $\mathcal{K}_2$ as sums of dilates of normalized bump functions. According to Corollary 6.2, we can assume that the flags $\mathcal{F}_1$ and $\mathcal{F}_2$ are given by

$$\mathcal{F}_1: \quad (0) \subseteq \mathbb{R}^{a_n} \subseteq \mathbb{R}^{a_{n-1}} \oplus \mathbb{R}^{a_n} \subseteq \cdots \subseteq \mathbb{R}^{a_2} \oplus \cdots \oplus \mathbb{R}^{a_n} \subseteq \mathbb{R}^N,$$

$$\mathcal{F}_2: \quad (0) \subseteq \mathbb{R}^{b_m} \subseteq \mathbb{R}^{b_{m-1}} \oplus \mathbb{R}^{b_m} \subseteq \cdots \subseteq \mathbb{R}^{b_2} \oplus \cdots \oplus \mathbb{R}^{b_m} \subseteq \mathbb{R}^N,$$

and the flag kernels are given by $\mathcal{K}_1 = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I$ and $\mathcal{K}_2 = \sum_{J \in \mathcal{E}_m} [\psi^J]_J$, where $\{\varphi^I \,|\, I \in \mathcal{E}_n\}$ is a uniformly bounded family of compactly supported functions with with common support and strong cancellation relative to the flag $\mathcal{F}_1$, and $\{\psi^J \,|\, J \in \mathcal{E}_m\}$ is a uniformly bounded family of compactly supported functions with with common support and strong cancellation relative to the flag $\mathcal{F}_2$.

If $\phi, \theta \in \mathcal{S}(\mathbb{R}^N)$, then

$$T_{\mathcal{K}_1}[\phi] = \lim_{F \nearrow \mathcal{E}_n} \sum_{I \in F} \phi * [\varphi^I]_I \quad \text{and} \quad T_{\mathcal{K}_2}[\theta] = \lim_{G \nearrow \mathcal{E}_m} \sum_{J \in G} \theta * [\psi^J]_J,$$

where the limits are in $L^2(\mathbb{R}^N)$ and are taken over finite subsets $F \subset \mathcal{E}_n$ and $G \subset \mathcal{E}_m$. For every fixed finite set $F \subset \mathcal{E}_n$, the function $\sum_{I \in F} \phi * [\varphi^I]_I \in \mathcal{S}(\mathbb{R}^N)$. Since $T_{\mathcal{K}_2}$ is a continuous mapping from $L^2(\mathbb{R}^N)$ to itself, it follows that

$$T_{\mathcal{K}_2}\big(T_{\mathcal{K}_1}[\phi]\big) = \lim_{F \nearrow \mathcal{E}_n} T_{\mathcal{K}_2}\Big(\sum_{I \in F} \phi * [\varphi^I]_I\Big) = \lim_{F \nearrow \mathcal{E}_n} \lim_{G \nearrow \mathcal{E}_m} \sum_{I \in F} \sum_{J \in G} \phi * [\varphi^I]_I * [\psi^J]_J$$

$$= \lim_{F \nearrow \mathcal{E}_n} \lim_{G \nearrow \mathcal{E}_m} \phi * \Big[\sum_{I \in F} \sum_{J \in G} [\varphi^I]_I * [\psi^J]_J \Big].$$

So in order to prove Theorem 7.3, we must study the finite sums $\sum_{I \in F} \sum_{J \in G} [\varphi^I]_I * [\psi^J]_J$, and show that these converge in the sense of distributions to a finite sum of flag kernels, each adapted to a flag which is equal to or coarser than $\mathcal{F}_0$.

Since the general situation is rather complicated, we first present an example which may help understand the difficulties.

7.3. An example

Suppose that we are working in $\mathbb{R}^5$ with the family of dilations given by

$$(7.2) \qquad \delta \cdot \mathbf{x} = \delta \cdot (x_1, x_2, x_3, x_4, x_5) = (\delta^{d_1} x_1, \delta^{d_2} x_2, \delta^{d_3} x_3, \delta^{d_4} x_4, \delta^{d_5} x_5)$$

with $d_1 \leq \cdots \leq d_5$. The standard flags on $\mathbb{R}^5$ correspond to partitions of $N = 5$.

Consider two flags $\mathcal{F}_1$ and $\mathcal{F}_2$ corresponding to the partitions $\mathfrak{A} = (2, 3)$, where we write $\mathbb{R}^5 = \mathbb{R}^2 \oplus \mathbb{R}^3$, and $\mathfrak{B} = (1, 2, 2)$, where we write $\mathbb{R}^5 = \mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$. Thus

$\mathcal{F}_1$ is the flag $(0) \subset \{x_1 = x_2 = 0\} \cong \mathbb{R}^3 \subset \mathbb{R}^5$,

$\mathcal{F}_2$ is the flag $(0) \subset \{x_1 = x_2 = x_3 = 0\} \cong \mathbb{R}^2 \subset \{x_1 = 0\} \cong \mathbb{R}^2 \oplus \mathbb{R}^2 \subset \mathbb{R}^5$.

We are given flag kernels

$$\mathcal{K}_1 = \sum_{I \in \mathcal{E}_2} [\varphi^I]_I \quad \text{and} \quad \mathcal{K}_2 = \sum_{J \in \mathcal{E}_3} [\psi^J]_J$$

adapted to these two flags. We then want to study the infinite sum

$$(7.3) \qquad \sum_{I \in \mathcal{E}_2} \sum_{J \in \mathcal{E}_3} [\varphi^I]_I * [\psi^J]_J$$

arising from the composition of the operators $T_{\mathcal{K}_1} \circ T_{\mathcal{K}_2}$.

Suppose $I = (i_1, i_2) \in \mathcal{E}_2$ and $J = (j_1, j_2, j_3) \in \mathcal{E}_3$, so that $i_1 \leq i_2$, and $j_1 \leq j_2 \leq j_3$. If $\varphi^I, \psi^J \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, we have

$$[\varphi^I]_I(\mathbf{x}) = 2^{-i_1(d_1+d_2)-i_2(d_3+d_4+d_5)}$$
$$\cdot \varphi^I \left(2^{-d_1 i_1} x_1, 2^{-d_2 i_1} x_2, 2^{-d_3 i_2} x_3, 2^{-d_4 i_2} x_4, 2^{-d_5 i_2} x_5 \right),$$
$$[\psi^J]_J(\mathbf{x}) = 2^{-j_1 d_1 - j_2(d_2+d_3) - j_3(d_4+d_5)}$$
$$\cdot \psi^J \left(2^{-d_1 j_1} x_1, 2^{-d_2 j_2} x_2, 2^{-d_3 j_2} x_3, 2^{-d_4 j_3} x_4, 2^{-d_5 j_3} x_5 \right).$$

Note that dilation by $I = (i_1, i_2)$ on $\mathbb{R}^2 \oplus \mathbb{R}^3$ is the same as dilation by the 5-tuple $\tilde{I} = (i_1, i_1, i_2, i_2, i_2)$ on $\mathbb{R}^5$, and dilation by $J = (j_1, j_2, j_3)$ on $\mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$ is the same as dilation by the 5-tuple $\tilde{J} = (j_1, j_2, j_2, j_3, j_3)$ on $\mathbb{R}^5$. Also note that we can reconstruct I and J from $\tilde{I}$ and $\tilde{J}$ by consolidating repeated indices. By Lemma 6.17, the convolution $[\varphi^I]_I * [\psi^J]_J$ is equal to $[\theta^{I,J}]_K$ where $\theta^{I,J} \in \mathcal{C}_0^\infty(\mathbb{R}^5)$, and where

$$(7.4) \qquad \begin{aligned} K &= (k_1, k_2, k_3, k_4, k_5) = \tilde{I} \vee \tilde{J} = (i_1, i_1, i_2, i_2, i_2) \vee (j_1, j_2, j_2, j_3, j_3) \\ &= (\max\{i_1, j_1\}, \max\{i_1, j_2\}, \max\{i_2, j_2\}, \max\{i_2, j_3\}, \max\{i_2, j_3\}). \end{aligned}$$

We must consider the sum in (7.3) of the convolutions $[\varphi^I]_I * [\psi^J]_J$, taken over all $I \in \mathcal{E}_2$ and $J \in \mathcal{E}_3$. Each pair $(I, J) \in \mathcal{E}_2 \times \mathcal{E}_3$ gives rise to a 5-tuple $K \in \mathcal{E}_5$. However, not all elements of $\mathcal{E}_5$ actually arise in this sum. (For example, it is clear from (7.4) that we must have $k_4 = k_5$, so the 5-tuple $(1, 2, 3, 4, 5)$ does not arise). Let $\mathcal{E}(\mathfrak{A}, \mathfrak{B})$ denote the set of all 5-tuples $K = (k_1, k_2, k_3, k_4, k_5)$ that do arise as in (7.4). (The notation reflects the fact that this set of 5-tuples is determined by the partitions $\mathfrak{A} = (2, 3)$ and $\mathfrak{B} = (1, 2, 2)$ of $N = 5$). Then for each $K \in \mathcal{E}(\mathfrak{A}, \mathfrak{B})$, let $\mathcal{E}(K)$ denote the set of pairs $(I, J) \in \mathcal{E}_2 \times \mathcal{E}_3$ which give rise to the 5-tuple K. Once $K \in \mathcal{E}(\mathfrak{A}, \mathfrak{B})$ is fixed, each of the terms in the inner infinite sum on the right-hand side of (7.3) is the K dilate of a normalized bump function $\theta^{I,J}$. Then we can write the sum in (7.3) as

$$(7.5) \quad \sum_{I \in \mathcal{E}_2} \sum_{J \in \mathcal{E}_3} [\varphi^I]_I * [\psi^J]_J = \sum_{K \in \mathcal{E}(\mathfrak{A}, \mathfrak{B})} \sum_{I, J \in \mathcal{E}(K)} [\varphi^I]_I * [\psi^J]_J = \sum_{K \in \mathcal{E}(\mathfrak{A}, \mathfrak{B})} \sum_{I, J \in \mathcal{E}(K)} [\theta^{I,J}]_K.$$

We will need to show that the infinite inner sum $\sum_{I,J\in\mathcal{E}(K)}[\theta^{I,J}]_K$ actually converges and is the K dilate of a normalized bump function. However, this is not enough to give the right description of the sum in (7.3) as a flag kernel. In the outer sum on the right-hand side of (7.5), the 5-tuple K runs over $\mathcal{E}(\mathfrak{A},\mathfrak{B})$ and *not* over all of $\mathcal{E}_5$. We still need to partition $\mathcal{E}(\mathfrak{A},\mathfrak{B})$ into a finite number of subsets based on which indices in K are repeated. To make this clear, we further analyze $\mathcal{E}(\mathfrak{A},\mathfrak{B})$.

As one sees from (7.4), the 5-tuple $K = \tilde{I} \vee \tilde{J}$ depends not only on the tuples $I = \{i_1, i_2\}$ and $J = \{j_1, j_2, j_3\}$, but also on the *ordering* of the larger set consisting of $\{i_1, i_2, j_1, j_2, j_3\}$. We know that $i_1 \leq i_2$ and $j_1 \leq j_2 \leq j_3$, but this does not determine the ordering of the larger set. Such orderings are in one-to-one correspondence with decompositions of the set $\{1, 2, 3, 4, 5\}$ into two disjoint subsets of sizes 2 and 3, where elements of the first set are indices from I, and elements of the second set are indices from J. Thus there are $\binom{5}{2} = \binom{5}{3} = 10$ such orderings. A description of these is given in the following Table 1:

Decomposition	Ordering	K	New decomposition	Free
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_1 = \{1,2\} \cup \{3,4,5\}$	$i_1 \leq i_2 \leq j_1 \leq j_2 \leq j_3$	$\{j_1, j_2, j_2, j_3, j_3\}$	$\mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$	i_1, i_2
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_2 = \{1,3\} \cup \{2,4,5\}$	$i_1 \leq j_1 < i_2 \leq j_2 \leq j_3$	$\{j_1, j_2, j_2, j_3, j_3\}$	$\mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$	i_1, i_2
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_3 = \{1,4\} \cup \{2,3,5\}$	$i_1 \leq j_1 \leq j_2 < i_2 \leq j_3$	$\{j_1, j_2, i_2, j_3, j_3\}$	$\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^2$	i_1
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_4 = \{1,5\} \cup \{2,3,4\}$	$i_1 \leq j_1 \leq j_2 \leq j_3 < i_2$	$\{j_1, j_2, i_2, i_2, i_2\}$	$\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^3$	i_1, j_3
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_5 = \{2,3\} \cup \{1,4,5\}$	$j_1 < i_1 \leq i_2 \leq j_2 \leq j_3$	$\{i_1, j_2, j_2, j_3, j_3\}$	$\mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$	i_2, j_1
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_6 = \{2,4\} \cup \{1,3,5\}$	$j_1 < i_1 \leq j_2 < i_2 \leq j_3$	$\{i_1, j_2, i_2, j_3, j_3\}$	$\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^2$	j_1
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_7 = \{2,5\} \cup \{1,3,4\}$	$j_1 < i_1 \leq j_2 \leq j_3 < i_2$	$\{i_1, j_2, i_2, i_2, i_2\}$	$\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^3$	j_1, j_3
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_8 = \{3,4\} \cup \{1,2,5\}$	$j_1 \leq j_2 < i_1 \leq i_2 \leq j_3$	$\{i_1, i_1, i_2, j_3, j_3\}$	$\mathbb{R}^2 \oplus \mathbb{R} \oplus \mathbb{R}^2$	j_1, j_2
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_9 = \{3,5\} \cup \{1,2,4\}$	$j_1 \leq j_2 < i_1 \leq j_3 < i_2$	$\{i_1, i_1, i_2, i_2, i_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	j_1, j_2, j_3
$\mathcal{E}(\mathfrak{A},\mathfrak{B})_{10} = \{4,5\} \cup \{1,2,3\}$	$j_1 \leq j_2 \leq j_3 < i_1 \leq i_2$	$\{i_1, i_1, i_2, i_2, i_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	j_1, j_2, j_3

TABLE 1. Decomposition of $\mathcal{E}(\mathfrak{A},\mathfrak{B})$.

In the first column, we have given the decomposition of $\{1, 2, 3, 4, 5\}$ into two subsets, the first with two elements and the second with three. This then gives an ordering of the elements in the set $\{i_1, i_2, i_3, i_4, i_5\}$ which is given in the second column. The third column gives the value of the 5-tuple $K = \tilde{I} \vee \tilde{J}$. In this tuple, certain entries can be repeated, and this corresponds to a new decomposition of $\mathbb{R}^5$. The fourth column gives this new decomposition of $\mathbb{R}^5$ dictated by the repeated indices of K. Finally, in each of the decompositions, certain of the indices from I or J appear in the 5-tuple K. In the sixth column of Table 1, we list the 'free'-variables which do *not* appear in K are listed. It is precisely these free variables which appear in the inner sum on the right-hand side of equation (7.5).

Table 1 shows that if $K \in \mathcal{E}(\mathfrak{A}, \mathfrak{B})$, then K takes one of five forms:

$$
\begin{array}{lll}
(k_1, k_1, k_2, k_2, k_2) & \text{(decomp. 9 and 10)} & \text{leading to the flag } \mathbb{R}^2 \oplus \mathbb{R}^3, \\
(k_1, k_2, k_2, k_3, k_3) & \text{(decomp. 1, 2 and 5)} & \text{leading to the flag } \mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2, \\
(k_1, k_2, k_3, k_3, k_3) & \text{(decomp. 4 and 7)} & \text{leading to the flag } \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^3, \\
(k_1, k_1, k_2, k_3, k_3) & \text{(decomp. 8)} & \text{leading to the flag } \mathbb{R}^2 \oplus \mathbb{R} \oplus \mathbb{R}^2, \\
(k_1, k_2, k_3, k_4, k_4) & \text{(decomp. 3 and 6)} & \text{leading to the flag } \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^2.
\end{array}
$$
(7.6)

The outer sum on the right-hand side of (7.5) thus splits into five separate sums:

$$
\begin{aligned}
\sum_{I \in \mathcal{E}_2} & \sum_{J \in \mathcal{E}_3} [\varphi^I]_I * [\psi^J]_J \\
&= \sum_{K \in \mathcal{E}_9 \cup \mathcal{E}_{10}} \sum_{I,J \in \widetilde{\mathcal{E}}(K)} [\varphi^I]_I * [\psi^J]_J + \sum_{K \in \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_5} \sum_{I,J \in \widetilde{\mathcal{E}}(K)} [\varphi^I]_I * [\psi^J]_J \\
&\quad + \sum_{K \in \mathcal{E}_4 \cup \mathcal{E}_7} \sum_{I,J \in \widetilde{\mathcal{E}}(K)} [\varphi^I]_I * [\psi^J]_J + \sum_{K \in E_8} \sum_{I,J \in \widetilde{\mathcal{E}}(K)} [\varphi^I]_I * [\psi^J]_J \\
&\quad + \sum_{K \in E_3 \cup E_6} \sum_{I,J \in \widetilde{\mathcal{E}}(K)} [\varphi^I]_I * [\psi^J]_J.
\end{aligned}
$$

Our object is to show that these five sums are flag kernels, each adapted to one of the five flags listed on the right-hand side of (7.6). To see this, we must show that in each case, the inner infinite sum converges, and has weak cancellation. Let us see why this happens in one case.

Case 1: $K \in E_9 \cup E_{10}$.

In this case, $K = \{i_1, i_1, i_2, i_2, i_2\}$ is fixed, and the inner sum $\sum_{I,J \in \widetilde{\mathcal{E}}(K)} [\varphi^I]_I *$ $[\psi^J]_J$ is over the free variables $\{j_1, j_2, j_3\}$ which satisfy the inequalities

$$
j_1 \leq j_2 < i_1 \leq j_3 < i_2 \qquad \text{or} \qquad j_1 \leq j_2 \leq j_3 < i_1 \leq i_2.
$$

In order to apply Theorem 6.8, we need to check that the sum converges to the K-dilate of a normalized bump function $\theta^{I,J}$, and moreover that $\theta^{I,J}$ has weak cancellation relative to the decomposition $\mathbb{R}^2 \oplus \mathbb{R}^3$.

To show that the sum over the free variables $\{j_1, j_2, j_3\}$ converges, we want to show that each term in the sum can be bounded by $2^{-\epsilon[(l_1 - j_1) + (l_2 - j_2) + (l_3 - j_3)]}$ where $l_1, l_2, l_3 \in \{i_2, i_2\}$, and $j_1 \leq l_1$, $j_2 \leq l_2$, and $j_3 \leq l_3$. This will follow because, by hypothesis, ψ^J has strong cancellation relative to the decomposition $\mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$. Thus ψ^J has cancellation in x_1, in either x_2 or x_3, and in either x_4 or x_5. In the variable x_1, $j_1 < i_1$, and so by Lemma 6.18, we get a gain of $2^{-\epsilon|i_1 - j_1|} \leq 2^{-\epsilon|i_1 - j_2|}$. If there is cancellation in x_2, we have $j_2 < i_1$, so we get a gain of $2^{-\epsilon|i_1 - j_2|}$, while if there is cancellation in x_3, we have $j_3 \leq i_2$ and so we get a gain of $2^{-\epsilon|i_2 - j_3|}$. Finally, if there is cancellation in x_4 or x_5, we have $j_3 < i_2$, and so we get a gain of $2^{-\epsilon|i_2 - j_3|}$. Taking the best of these estimates, we see that the size of $[\varphi^I]_I * [\psi^J]_J$ is dominated by a constant times $2^{-\epsilon[|i_1 - j_1| + |i_1 - j_2| + |i_2 - j_3|]}$. Thus in this case we can take $l_1 = i_1$, $l_2 = i_1$, and $l_3 = i_2$.

The key points in this convergence argument are the following:

(a) If f_s is a free variable, it does not appear in K. Since the entries $k_s \in K$ are the maxima of the corresponding entries of $i_s \in \tilde{I}$ and $j_s \in \tilde{J}$, the free variable must satisfy $f_s \leq k_s$.

(b) Since we will sum over the free variable f_s, but the variable $k_s \in K$ is fixed, it suffices to show that there is a gain $2^{-\epsilon(k_s - f_s)}$.

(c) The function φ^I or ψ^J with the free variable f_s may not necessarily have cancellation in the variable x_s. (For example, the free variable j_2 comes from the function $\psi^{(j_1, j_2, j_2, j_3, j_3)}$, and we only know that this function has cancellation in the variable x_2 *or* the variable x_3). However, if there is no cancellation in the free variable, there is a *smaller* free variable where there is cancellation, and where the corresponding element of K is the same. (In our example, $\psi^{(j_2, j_2, j_3)}$ has cancellation in x_1, and $k_1 = k_2$).

To see that the sum of the terms $[\varphi^I]_I * [\psi^J]_J$ has weak cancellation relative to the decomposition $\mathbb{R}^2 \oplus \mathbb{R}^3$, we again use Lemma 6.18. We only need to observe that either φ^I or ψ^J has cancellation in one of the variables $\{x_1, x_2\}$, and also that either φ^I or ψ^J has cancellation in one of the variables $\{x_3, x_4, x_5\}$. But this is clear: for example, ψ^J has strong cancellation relative to the decomposition $\mathbb{R} \oplus \mathbb{R}^2 \oplus \mathbb{R}^2$, and so has cancellation in x_1, and φ^I has strong cancellation relative to the decomposition $\mathbb{R}^2 \oplus \mathbb{R}^3$, and so has cancellation in one of the variables $\{x_3, x_4, x_5\}$.

7.4. The general decomposition

Now let us return to the general situation. Suppose $\mathcal{F}_1$ and $\mathcal{F}_2$ are standard flags arising from two (in general different) decompositions we label as $(\mathfrak{A})$ and $(\mathfrak{B})$:

$$(\mathfrak{A}): \quad \mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n},$$
$$(\mathfrak{B}): \quad \mathbb{R}^N = \mathbb{R}^{b_1} \oplus \cdots \oplus \mathbb{R}^{b_m}.$$

Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be flag kernels adapted to the flags $\mathcal{F}_1$ and $\mathcal{F}_2$. We only need to consider the parts of these kernels given by sums of dilates of normalized bump functions with strong cancellation. (That is, for each kernel we focus on the part called $\mathcal{K}_0$ in Theorem 6.1 and disregard the other terms since they correspond to coarser flags). Thus we can write

$$(7.7) \qquad \mathcal{K}_1 = \sum_{I \in \mathcal{E}_n} [\varphi^I]_I, \quad \mathcal{K}_2 = \sum_{J \in \mathcal{E}_m} [\psi^J]_J,$$

where each φ^I has strong cancellation relative to the decomposition $\mathfrak{A}$ and each ψ^J has strong cancellation relative to the decomposition $\mathfrak{B}$. Let

$$\mathcal{K}_1^F = \sum_{I \in F \subset \mathcal{E}_n} [\varphi^I]_I, \quad \mathcal{K}_2^G = \sum_{J \in G \subset \mathcal{E}_m} [\psi^J]_J,$$

where $F \subset \mathcal{E}_n$ and $G \subset \mathcal{E}_m$ are finite subsets. We study the double sum

$$(7.8) \qquad \mathcal{K}_1^F * \mathcal{K}_2^G = \sum_{I \in F \subset \mathcal{E}_n} \sum_{J \in G \subset \mathcal{E}_m} [\varphi^I]_I * [\psi^J]_J.$$

Let $\mathcal{F}_0$ be the coarsest flag which is finer than both $\mathcal{F}_1$ and F_2.

Theorem 7.4. *Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be the flag kernels given in* (7.7). *Then*

$$\lim_{\substack{F \nearrow \mathcal{E}_n \\ G \nearrow \mathcal{E}_m}} \mathcal{K}_1^F * \mathcal{K}_2^G = \mathcal{K}_1 * \mathcal{K}_2$$

converges in the sense of distributions to a finite sum of flag kernels $\sum \mathcal{K}_\mu$, each of which is adapted to a flag $\mathcal{F}_\mu$ which is equal to or coarser than the flag $\mathcal{F}_0$.

Before outlining the proof, we review our notation. If $\mathbf{x} \in \mathbb{R}^N$, we can write $\mathbf{x} = (\mathbf{x}_1', \ldots, \mathbf{x}_n')$ with $\mathbf{x}_j' = (x_{p_j'}, \ldots, x_{q_j'}) \in \mathbb{R}^{a_j}$, or $\mathbf{x} = (\mathbf{x}_1'', \ldots, \mathbf{x}_m'')$ with $\mathbf{x}_k'' = (x_{p_k''}, \ldots, x_{q_k''}) \in \mathbb{R}^{b_k}$. We let $J_j' = \{p_j', \ldots, q_j'\}$ and $J_k'' = \{p_k'', \ldots, q_k''\}$ so that $\{1, \ldots, N\} = \bigcup_{j=1}^n J_j' = \bigcup_{k=1}^m J_k''$. Define

$$\sigma_{\mathfrak{A}} : \{1, \ldots, N\} \to \{1, \ldots, n\} \quad \text{so that} \quad l \in J_{\sigma_{\mathfrak{A}}(l)}' \text{ for } 1 \leq l \leq N;$$
$$\tau_{\mathfrak{B}} : \{1, \ldots, N\} \to \{1, \ldots, m\} \quad \text{so that} \quad l \in J_{\sigma_{\mathfrak{B}}(l)}'' \text{ for } 1 \leq l \leq N.$$

If Q_j' is the homogeneous dimension of $\mathbb{R}^{a_j}$ and Q_k'' is the homogeneous dimension of $\mathbb{R}^{b_k}$, then

$$Q_j' = d_{p_j'} + \cdots + d_{q_j'} = \sum_{l \in J_j'} d_l, \quad Q_k'' = d_{p_k''} + \cdots + d_{q_k''} = \sum_{l \in J_k''} d_l.$$

If $I \in \mathcal{E}_n$ and $J \in \mathcal{E}_m$, the notation $[\varphi^I]_I$ and $[\psi^J]_J$ refers to the families of dilations

$$[\varphi^I](\mathbf{x}) = 2^{-[Q_1' i_1 + \cdots + Q_n' i_n]} \varphi(2^{-i_1} \cdot \mathbf{x}_1', \ldots, 2^{-i_n} \cdot \mathbf{x}_n'),$$
$$[\psi^J](\mathbf{x}) = 2^{-[Q_1'' j_1 + \cdots + Q_m'' j_m]} \varphi(2^{-j_1} \cdot \mathbf{x}_1'', \ldots, 2^{-j_m} \cdot \mathbf{x}_m'').$$

In order to compare multi-indices $I = (i_1, \ldots, i_n) \in \mathcal{E}_n$ and $J = (j_1, \ldots, j_m) \in \mathcal{E}_m$ which parameterize different families of dilations, we identify them with multi-indices of length N with repeated entries. Thus we define $p_{\mathfrak{A}} : \mathcal{E}_n \to E_N$ and $p_{\mathfrak{B}} : \mathcal{E}_m \to E_N$ so that $p_{\mathfrak{A}}(I)$ is the N-tuple with i_1 repeated a_1 times, i_2 repeated a_2 times, *etc.* We define $p_{\mathfrak{B}}$ analogously. Thus

$$p_{\mathfrak{A}}(I) = I = (I_1, \ldots, I_N) = (\overbrace{i_1, \ldots \ldots \ldots, i_1}^{a_1}, \ldots, \overbrace{i_r, \ldots, i_r}^{a_r}, \ldots \ldots \ldots, \overbrace{i_n, \ldots, i_n}^{a_n}),$$

$$p_{\mathfrak{B}}(J) = J = (J_1, \ldots, J_N) = (\overbrace{j_1, \ldots, j_1}^{b_1}, \ldots \ldots, \overbrace{j_s, \ldots \ldots, j_s}^{b_s}, \ldots, \overbrace{j_m, \ldots \ldots, j_m}^{b_m}).$$

Explicitly, $p_{\mathfrak{A}}(I) = (I_1, \ldots, I_N)$ and $p_{\mathfrak{B}}(J) = (J_1, \ldots, J_N)$ where $I_l = i_{\sigma(l)}$ and $J_l = j_{\tau(l)}$. Next set $K = (K_1, \ldots, K_N) = p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) = (k_1, \ldots, k_N)$. This means that for $1 \leq l \leq N$

$$(7.9) \qquad K_l = i_{\sigma(l)} \vee j_{\tau(l)} = \max\{i_{\sigma(l)}, j_{\tau(l)}\} = \begin{cases} j_{\tau(l)} & \text{if } i_{\sigma(l)} \leq j_{\tau(l)} \\ i_{\sigma(l)} & \text{if } j_{\tau(l)} < i_{\sigma(l)} \end{cases}.$$

Note that we can then define a map $\pi_{I,J} : \{1, \ldots, N\} \to \{1, \ldots, n, n+1, \ldots, n+m\}$ so that

$$(7.10) \qquad \pi_{I,J}(l) = \begin{cases} \ell \in \{1, \ldots n\} & \text{if } K_l = i_{\sigma(\ell)}; \\ n + \ell \in \{n+1, \ldots, n+m\} & \text{if } K_l = j_{\tau(\ell)}. \end{cases}$$

Now let us outline the proof of Theorem 7.4. For each $I \in \mathcal{E}_n$ and $J \in \mathcal{E}_m$ it follows from Lemma 6.17 that there is a function $\theta^{I,J} \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, normalized relative to φ^I and ψ^J, so that

$$(7.11) \qquad [\varphi^I]_I * [\psi^J]_J = [\theta^{I,J}]_{p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J)},$$

and hence equation (7.8) can be written

$$(7.12) \qquad \mathcal{K}_1^F * \mathcal{K}_2^G = \sum_{(I,J) \in F \times G \subset \mathcal{E}_n \times \mathcal{E}_m} [\theta^{I,J}]_{p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J)}.$$

We analyze this sum by decomposing the set $\mathcal{E}_n \times \mathcal{E}_m$ into disjoint subsets. Let $\mathfrak{P}(n, m)$ denote the set of permutations $\mu : \{1, \ldots, n, n+1, \ldots, n+m\} \to \{1, \ldots, n, n+1, \ldots, n+m\}$ which preserve the order of the first n elements $\{1, \ldots, n\}$ and of the last m elements $\{n+1, \ldots, n+m\}$. (Explicitly, this means that if $\mu \in \mathfrak{P}(n, m)$, then $1 \leq s < t \leq n$ implies $\mu(s) < \mu(t)$ and $n + 1 \leq s < t \leq n + m$ implies $\mu(s) < \mu(t)$). This corresponds to the ten cases in the example studied in Section 7.3. The cardinality of $\mathfrak{P}(n, m)$ is $\binom{n+m}{n}$. Let $I \in \mathcal{E}_n$ and $J \in \mathcal{E}_m$, and let us write

$$I = (\alpha_1, \ldots, \alpha_n) \quad \text{and} \quad J = (\alpha_{n+1}, \ldots, \alpha_{n+m}), \quad \text{so that}$$
$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_n \quad \text{and} \quad \alpha_{n+1} \leq \alpha_{n+2} \leq \cdots \leq \alpha_{n+m}.$$

Then let $L(I, J)$ be the (weakly) increasing rearrangement of the set $I \cup J = \{\alpha_1, \ldots, \alpha_{n+m}\}$ so that
(7.13)

if $1 \leq r < s \leq n$	then α_r comes to the left of α_s,
if $n + 1 \leq r < s \leq n + m$	then α_r comes to the left of α_s,
if $1 \leq r \leq n, n+1 \leq s \leq n+m$ and $\alpha_r < \alpha_s$	then α_r comes to the left of α_s,
if $1 \leq r \leq n, n+1 \leq s \leq n+m$ and $\alpha_s \leq \alpha_r$	then α_s comes to the left of α_r.

This rearrangement of $\{\alpha_1, \ldots, \alpha_{n+m}\}$ is given by $L(I, J) = \{\alpha_{\mu(1)}, \ldots, \alpha_{\mu(n+m)}\}$ where μ is a permutation of the set of subscripts $\{1, \ldots, n, n + 1, \ldots, n + m\}$, and it follows from (7.13) that $\mu \in \mathfrak{P}(n, m)$. In this way we associate to each pair $(I, J) \in \mathcal{E}_n \times \mathcal{E}_m$ a unique $\mu = \mu(I, J) \in \mathfrak{P}(n, m)$. Conversely, for each $\mu \in \mathfrak{P}(n, m)$, let

$$(7.14) \qquad \mathcal{E}_{n,m}(\mu) = \big\{ (I, J) \in \mathcal{E}_n \times \mathcal{E}_m \,\big|\, L(I, J) = \{\alpha_{\mu(1)}, \ldots, \alpha_{\mu(n+m)}\} \big\}.$$

It follows that we have a disjoint decomposition $\mathcal{E}_n \times \mathcal{E}_m = \bigcup_{\mu \in \mathfrak{P}(m,n)} \mathcal{E}_{n,m}(\mu)$, and we can write equation (7.8) as

$$(7.15) \qquad \mathcal{K}_1^F * \mathcal{K}_2^G = \sum_{\mu \in \mathfrak{P}(m,n)} \bigg(\sum_{(I,J) \in \mathcal{E}_{n,m}(\mu) \cap (F \times G)} [\theta^{I,J}]_{p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J)} \bigg).$$

Now let

$$(7.16) \qquad E_N(\mu) = \left\{ K \in E_N \mid K = p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) \text{ where } (I,J) \in \mathcal{E}_{n,m}(\mu) \right\}.$$

Note that in general $E_N(\mu)$ is a proper subset of E_N, and an element $K \in E_N(\mu)$ can be represented in many ways as $p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J)$ with $(I,J) \in \mathcal{E}_{n,m}(\mu)$. We can then rewrite (7.15) as

$$
\begin{aligned}
(7.17) \qquad
\mathcal{K}_1^F * \mathcal{K}_2^G &= \sum_{\mu \in \mathfrak{P}(m,n)} \left(\sum_{K \in E_N(\mu)} \left(\sum_{\substack{(I,J) \in \mathcal{E}_{n,m}(\mu) \cap (F \times G) \\ p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) = K}} [\theta^{I,J}]_K \right) \right) \\
&= \sum_{\mu \in \mathfrak{P}(m,n)} \left(\sum_{K \in E_N(\mu)} \left[\sum_{\substack{(I,J) \in \mathcal{E}_{n,m}(\mu) \cap (F \times G) \\ p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) = K}} \theta^{I,J} \right]_K \right)
\end{aligned}
$$

We will prove in Lemma 7.8 below that because the functions $\{\varphi^I\}$ and $\{\psi^J\}$ have strong cancellation, the innermost sum

$$(7.18) \qquad \sum_{\substack{(I,J) \in \mathcal{E}_{n,m}(\mu) \\ p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) = K}} \theta^{I,J} = \lim_{\substack{F \nearrow \mathcal{E}_n \\ G \nearrow \mathcal{E}_m}} \sum_{\substack{(I,J) \in \mathcal{E}_{n,m}(\mu) \cap (F \times G) \\ p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) = K}} \theta^{I,J}$$

converges to a function $\Theta^K \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ which is normalized relative to the families $\{\varphi^I\}$ and $\{\psi^j\}$. From this it follows from (7.17) that

$$(7.19) \qquad \mathcal{K}_1 * \mathcal{K}_2 = \lim_{\substack{F \nearrow \mathcal{E}_n \\ G \nearrow \mathcal{E}_m}} \mathcal{K}_1^F * \mathcal{K}_2^G = \sum_{\mu \in \mathfrak{P}(n,m)} \left(\sum_{K \in E_N(\mu)} [\Theta^K]_K \right).$$

We will also see in Lemma 7.8 below that for each fixed μ, the functions $\{\Theta^K\}$ for $K \in E_N(\mu)$ have weak cancellation relative to a decomposition of $\mathbb{R}^N$ depending on μ, $\mathbb{R}^N = \mathbb{R}^{c_1} \oplus \cdots \oplus \mathbb{R}^{c_r}$, and hence the inner sum on the right hand side of (7.19) is a flag kernel relative to the corresponding standard flag $\mathcal{F}_\mu$. Once this is done, we will have established Theorem 7.4.

We now turn to the details of the proof. We begin by studying the N-tuple $K = (K_1, \ldots, K_N) = p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J)$ if $(I,J) \in \mathcal{E}_{n,m}(\mu)$. Partition K into disjoint subsets of consecutive entries where two successive elements K_l and K_{l+1} belong to the same subset if and only if either

(i) $K_l = i_{\sigma(l)} \geq j_{\tau(l)}$, $K_{l+1} = i_{\sigma(l+1)} \geq j_{\tau(l+1)}$, and $\sigma(l) = \sigma(l+1)$;   or

(ii) $K_l = j_{\tau(l)} \geq i_{\sigma(l)}$, $K_{l+1} = j_{\tau(l+1)} \geq i_{\sigma(l+1)}$, and $\tau(l) = \tau(l+1)$.

In particular, if two successive elements K_l and K_{l+1} belong to the same subset, they must be equal. Thus we write

$$K = (K_1, \ldots, K_N) = \left(\{K_{\alpha_1}, \ldots, K_{\beta_1}\}, \{K_{\alpha_2}, \ldots, K_{\beta_2}\}, \ldots, \{K_{\alpha_r}, \ldots, K_{\beta_r}\} \right)$$

where

$$K_{\alpha_1} = K_{\alpha_1+1} = \cdots = K_{\beta_1},$$
$$K_{\alpha_2} = K_{\alpha_2+1} = \cdots = K_{\beta_2},$$
$$\vdots$$
$$K_{\alpha_r} = K_{\alpha_r+1} = \cdots = K_{\beta_r}.$$

We can also write

$$(7.20) \qquad K = (K_1, \ldots, K_N) = \big(\overbrace{k_1, \ldots, k_1}^{c_1} ; \overbrace{k_2, \ldots, k_2}^{c_2} ; \ldots ; \overbrace{k_r, \ldots, k_r}^{c_r} \big),$$

so that $c_1 + c_2 + \cdots + c_r = N$. Note that $1 \le r < m+n$ since either i_1 or j_1 does not appear in K. We have the following properties of this decomposition:

Proposition 7.5. *The integers $\{\alpha_1, \beta_1, \ldots, \alpha_r, \beta_r\}$ depend only on the permutation μ and are independent of the choice of $(I, J) \in \mathcal{E}_{n,m}(\mu)$. In fact, let $\widetilde{K}_l = \{K_{\alpha_l}, \ldots, K_{\beta_l}\}$ be one of the subsets of consecutive indices in $K = p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) \in \mathcal{E}_{n,m}(\mu)$. Then:*

(1) *The starting position α_l coincides either with the starting position of one of the subsets of $p_{\mathfrak{A}}(I)$ or with the position of one of the subsets of $p_{\mathfrak{B}}(J)$. More precisely, for $l \ge 2$,*

    (1a) *If $K_{\alpha_l} = i_{\sigma(\alpha_l)}$, then $\sigma(\alpha_l) > \sigma(\alpha_l - 1)$, so the start of $\widetilde{K}_l$ coincides with the start of the index $i_{\sigma(\alpha_l)}$ in $p_{\mathfrak{A}}(I)$.*

    (1b) *If $K_{\alpha_l} = j_{\tau(\alpha_l)}$, then $\tau(\alpha_l) > \tau(\alpha_l - 1)$, so the start of $\widetilde{K}_l$ coincides with the start of the index $j_{\tau(\alpha_l)}$ in $p_{\mathfrak{B}}(J)$.*

(2) *If the ending position of $\widetilde{K}_l$ does not coincide with the end of the corresponding segment of $p_{\mathfrak{A}}(I)$ or $p_{\mathfrak{B}}(J)$, the entries of the segment which do not appear in $\widetilde{K}_l$ are bounded above by the corresponding entries of K. More precisely, for $\gamma \le N$,*

    (2a) *Suppose that $K_{\alpha_l} = i_{\sigma(\alpha_l)}$, so that $K_t = i_{\sigma(\alpha_l)}$ for $\alpha_l \le t \le \beta_l$. Suppose that $I_t = I_{\alpha_l}$ for $\alpha_l \le t \le \gamma$ and that $\gamma > \beta_l$. Then $I_t \le J_t = K_t$ for $\beta_l + 1 \le t \le \gamma$.*

    (2b) *Suppose that $K_{\alpha_l} = j_{\tau(\alpha_l)}$, so that $K_t = j_{\tau(\alpha_l)}$ for $\alpha_l \le t \le \beta_l$. Suppose that $J_t = J_{\alpha_l}$ for $\alpha_l \le t \le \gamma$ and that $\gamma > \beta_l$. Then $J_t \le I_t = K_t$ for $\beta_l + 1 \le t \le \gamma$.*

Proof. We begin by establishing part (1a). If it were not true, then since $j_{\tau(\alpha_l-1)} \le j_{\tau(\alpha_l)}$, we would have

$$j_{\tau(\alpha_l-1)} \le j_{\tau(\alpha_l)} \le i_{\sigma(\alpha_l)} = i_{\sigma(\alpha_l-1)},$$

and by condition (i) in the definition of $\widetilde{K}_l$ it would follow that K_{α_l} and K_{α_l-1} belong to the same subset. Part (1b) follows in the same way. To establish part (2a), observe that since $J_{\beta_l+1} \le J_t$, it suffices to show this for $t = \beta_l + 1$. However, if $J_{\beta_l+1} < I_{\beta_l+1}$, it would follow from (i) that K_{β_l+1} belongs to the same subset as K_{α_l}. Part (2b) follows in the same way.

Now parts (1) and (2) show that the positions where we decompose K depend on the *ordering* of the entries of $p_{\mathfrak{A}}(I)$ and $p_{\mathfrak{B}}(J)$, and not on the entries themselves. This shows that the decomposition depends only on μ, completing the proof.   □

Definition 7.6. Let $\mu \in \mathfrak{P}(n,m)$.

(1) It follows from Proposition 7.5 that the permutation μ determines r and the integers $\{c_1, \ldots, c_r\}$. Thus μ determines the decomposition $\mathbb{R}^N = \mathbb{R}^{c_1} \oplus \cdots \oplus \mathbb{R}^{c_r}$. We let $\mathcal{F}_\mu$ denote the corresponding standard flag $(0) \subseteq \mathbb{R}^{c_r} \subseteq \mathbb{R}^{c_{r-1}} \oplus \mathbb{R}^{c_r} \subseteq \cdots \subseteq \mathbb{R}^{c_2} \oplus \cdots \oplus \mathbb{R}^{c_n} \subseteq \mathbb{R}^{c_1} \oplus \cdots \oplus \mathbb{R}^{c_r} = \mathbb{R}^N$. For $\mathbf{x} \in \mathbb{R}^N$, we write $\mathbf{x} = (\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_r)$, and we let $\{\widetilde{J}_1, \ldots, \widetilde{J}_r\}$ denote the corresponding sets of subscripts so that x_l is a coordinate in $\mathbb{R}^{c_k}$ if and only if $l \in \widetilde{J}_k$.

(2) Let $I = (i_1, \ldots, i_n), J = (j_1, \ldots, j_m) \in \mathcal{E}_{n,m}(\mu)$. An index i_l is *free* if $I_t \leq J_t$ for all t such that $\sigma(t) = l$. An index j_l is *free* if $J_t < I_t$ for all t such that $\sigma(t) = l$. In particular, a free index does *not* appear in the set $K = p_{\mathfrak{A}}(I) \vee p_{\mathfrak{B}}(J) \in E_N(\mu)$.

Note that whether or not an index is free depends only on the choice of μ, and not on the choice of $(I,J) \in \mathcal{E}_{n,m}(\mu)$. The number of free elements is equal to $m + n - r$, and $1 \leq m + n - r < m + r$.

Proposition 7.7. *Fix $\mu \in \mathfrak{P}(n,m)$, and let $(I,J) \in \mathcal{E}_{n,m}(\mu)$ with $I = (i_1, \ldots, i_n)$ and $J = (j_1, \ldots, j_m)$. Let $\varphi^I, \psi^J \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, and suppose that φ^I has strong cancellation relative to the decomposition $\mathbb{R}^N = \mathbb{R}^{a_1} \oplus \cdots \oplus \mathbb{R}^{a_n}$, and that ψ^J has strong cancellation relative to the decomposition $\mathbb{R}^N = \mathbb{R}^{b_1} \oplus \cdots \oplus \mathbb{R}^{b_m}$.*

(1) *Suppose that an index i_l is free, and that $i_l = I_r = I_{r+1} = \cdots = I_s$ is the corresponding group of indices in $p_{\mathfrak{A}}(I)$, so that $I_{r-1} = i_{l-1}$ and $I_{s+1} = i_{l+1}$. Then the function φ^I has cancellation in the variables $\{x_{I_r}, \ldots, x_{I_s}\}$, and $I_t \leq J_t$ for $r \leq t \leq s$.*[8]

(2) *Suppose that an index j_l is free, and that $\{j_l = J_r = J_{r+1} = \cdots = J_s\}$ is the corresponding group of indices in $p_{\mathfrak{B}}(J)$, so that $J_{r-1} = j_{l-1}$ and $J_{s+1} = j_{l+1}$. Then the function ψ^J has cancellation with respect to one of the variables $\{x_{J_r}, \ldots, x_{J_s}\}$, and $J_t \leq I_t$ for $r \leq t \leq s$.*[9]

(3) *Let $\mathbb{R}^N = \mathbb{R}^{c_1} \oplus \cdots \oplus \mathbb{R}^{c_r}$ be the decomposition corresponding to μ, and let $\{\widetilde{J}_1, \ldots, \widetilde{J}_r\}$ be the corresponding sets of subscripts. Let $1 \leq l \leq r$.*

   (a) *Suppose that $K_{\alpha_l} = \cdots = K_{\beta_l} = i_\ell$ so that $\sigma(\alpha_l) = \cdots = \sigma(\beta_l) = \ell$. Then either φ^I has cancellation in a variable x_t with $t \in \widetilde{J}_l$, or φ^I has cancellation in a coordinate x_t with $t > \beta_l$, in which case $K_t = J_t \geq I_t = i_\ell$.*

   (b) *Suppose that $K_{\alpha_l} = \cdots = K_{\beta_l} = j_\ell$ so that $\tau(\alpha_l) = \cdots = \tau(\beta_l) = \ell$. Then either ψ^J has cancellation in a variable x_t with $t \in \widetilde{J}_l$, or ψ^J has cancellation in a coordinate x_t with $t > \beta_l$, in which case $K_t = I_t \geq J_t = j_\ell$.*

---

[8]It follows from Lemma 5.1 that φ^I can be written as a sum of functions each of which has cancellation in one of the variables $\{x_{I_r}, \ldots, x_{I_s}\}$.

[9]It follows from Lemma 5.1 that ψ^J can be written as a sum of functions each of which has cancellation in one of the variables $\{x_{J_r}, \ldots, x_{J_s}\}$.

Proof. To prove assertion (1), note that since i_l is free, it does not appear in K. Hence for any $r \leq t \leq s$, $i_l \neq \max\{I_t, J_t\} = \max\{i_l, J_t\}$, and so $J_\ell \geq i_\ell = I_t$. Since φ^I is assumed to have strong cancellation and $\{x_{I_r}, \ldots, x_{I_s}\}$ are precisely the variables corresponding to the index i_r, this establishes (1). The proof of assertion (2) proceeds in the same way.

To prove assertion (3a), note that by Proposition 7.5, part (1a), the only way in which it is possible for φ^I not to have cancellation in a variable x_t with $t \in \tilde{J}_l$ is if $I_t = I_{\alpha_l}$ for $\alpha_l \leq t \leq \gamma$ and that $\gamma > \beta_l$. But then the conclusion follows from Proposition 7.5, part (2a). The proof of assertion (3b) follows in the same way. □

Lemma 7.8. *Fix $\mu \in \mathfrak{P}(n, m)$.*

(1) *Let $K \in E_N(\mu)$. Then there exists $\Theta^K \in \mathcal{C}_0^\infty(\mathbb{R}^N)$, normalized relative to the families $\{\varphi^I\}$ and $\{\psi^J\}$ so that the sum*

$$\sum_{\substack{(I,J) \in \mathcal{E}_{n,m}(\mu) \\ p_\mathfrak{A}(I) \vee p_\mathfrak{B}(J) = K}} [\varphi^I]_I * [\psi^J]_J$$

converges (uniformly) to $[\Theta^K]_K$.

(2) *The function Θ^K has weak cancellation relative to the decomposition of $\mathbb{R}^N$ corresponding to μ.*

Proof. Let $(I, J) \in \mathcal{E}_{n,m}(\mu)$. Suppose that φ^I has cancellation in the variables $\{x_{r_1}, \ldots, x_{r_n}\}$, with $r_l \in J_l'$, and that ψ^J has cancellation in the variables $\{x_{s_1}, \ldots, x_{s_m}\}$ with $s_l \in J_l''$. Let

$$A_0 = \{l \in \{1, \ldots, n\} \,|\, i_{r_l} < j_{r_l}\}, \qquad B_0 = \{l \in \{1, \ldots, m\} \,|\, j_{s_l} < i_{s_l}\}.$$

Then it follows from Proposition 7.7 that if i_l is a free index, $l \in A_0$ and if j_l is a free index, then $l \in B_0$. On the other hand, according to Lemma 6.18, we can write each $[\varphi^I]_I * [\psi^J]_J$ as a finite sum of terms of the form

$$\prod_{s \in A'} 2^{-\epsilon|j_{l_s} - i_{l_s}|} \prod_{t \in B'} 2^{-\epsilon|i_{m_t} - j_{m_t}|} \prod_{s \in A''} 2^{-\epsilon|i_{l_s+1} - i_{l_s}|}$$

$$(7.21) \qquad\qquad \prod_{t \in B''} 2^{-\epsilon|j_{m_t+1} - j_{m_t}|} \prod_{s \in A'''} \partial_{l_s} \prod_{t \in B'''} \partial_{m_t} [\tilde{\theta}]$$

where

1. A', A'', A''' are disjoint subsets of $\{1, \ldots, n\}$ with $A' \cup A'' \cup A''' = \{1, \ldots, n\}$;
2. B', B'', B''' are disjoint subsets of $\{1, \ldots, m\}$ with $B' \cup B'' \cup B''' = \{1, \ldots, m\}$;
3. $A' \subset A_0 \subset A' \cup A''$ and $B' \subset B_0 \subset B' \cup B''$, and $n \notin A''$ and $n \notin B''$;
4. each function $\tilde{\theta}$ depends on $\{A', A'', A''', B', B'', B'''\}$ and is normalized relative to φ and ψ.

Since K is fixed, the sum for $p_\mathfrak{A}(I) \vee p_\mathfrak{B}(J) = K$ is precisely the sum over the set of free indices in $\{i_1, \ldots, i_n, j_1, \ldots, j_m\}$, and these are contained in the indices in $A' \cup A'' \cup B' \cup B''$. The exponential decay in the powers of 2 in equation (7.21) show that the sum over all the free indices converges, and what remains satisfies the requirements for weak cancellation in Definition 5.5. This completes the proof. □

7.5. Further examples

It may help to consider two additional examples.

Example 2: Suppose that $N = 5$, and that the two partitions of $\{1, 2, 3, 4, 5\}$ are $A = \{2, 3\}$ and $B = \{2, 3\}$. Thus A and B come from the same decomposition $\mathbb{R}^5 = \mathbb{R}^2 \oplus \mathbb{R}^3$. There are $\binom{4}{2} = 6$ different decompositions of $\{1, 2, 3, 4\}$ into two disjoint subsets of cardinalities 2 and 2. The six decompositions, the 5-tuples $\tilde{I}$, $\tilde{J}$, and $\tilde{I} \vee \tilde{J}$, and the resulting new decomposition C of $\{1, 2, 3, 4, 5\}$ are listed in Table 2 below. This example is typical of the convolution of two kernels coming from the *same* flag (in this case coming from the decomposition $\{2, 3\}$). Note that all decompositions lead to the same decomposition $\{2, 3\}$. Thus the convolution will be a flag of the same type.

Decomposition	Ordering	K	New decomposition	C	*Free variables*
$\{1, 2\} \cup \{3, 4\}$	$i_1 \le i_2 \le j_1 \le j_2$	$\{j_1, j_1, j_2, j_2, j_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	$\{2, 3\}$	i_1, i_2
$\{1, 3\} \cup \{2, 4\}$	$i_1 \le j_1 < i_2 \le j_2$	$\{j_1, j_1, j_2, j_2, j_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	$\{2, 3\}$	i_1, i_2
$\{1, 4\} \cup \{2, 3\}$	$i_1 \le j_1 \le j_2 < i_2$	$\{j_1, j_1, i_2, i_2, i_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	$\{2, 3\}$	i_1, j_2
$\{2, 3\} \cup \{1, 4\}$	$j_1 < i_1 \le i_2 \le j_2$	$\{i_1, i_1, j_2, j_2, j_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	$\{2, 3\}$	i_2, j_1
$\{2, 4\} \cup \{1, 3\}$	$j_1 < i_1 \le j_2 < i_2$	$\{i_1, i_1, i_2, i_2, i_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	$\{2, 3\}$	j_1, j_2
$\{3, 4\} \cup \{1, 2\}$	$j_1 \le j_2 < i_1 \le i_2$	$\{i_1, i_1, i_2, i_2, i_2\}$	$\mathbb{R}^2 \oplus \mathbb{R}^3$	$\{2, 3\}$	j_1, j_2

TABLE 2.

Example 3: Suppose that $N = 5$, and that the two partitions of $\{1, 2, 3, 4, 5\}$ are $A = \{5\}$ and $B = \{1, 2, 3, 4, 5\}$. Thus $m = 1$ and $n = 5$. There are then $\binom{6}{5} = 6$ different decompositions of $\{1, 2, 3, 4, 5, 6\}$ into two disjoint subsets of cardinalities 1 and 5. The six decompositions, the 5-tuples $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{I} \vee \mathcal{J}$, and the resulting new decomposition C of $\{1, 2, 3, 4, 5\}$ are listed in Table 3 below.

This example is typical of the convolution of a Calderón–Zygmund kernel with a kernel that is as fine as possible. In this case, the flag A is coarser than flag B.

Decomposition	Ordering	K	New decomposition	C	Free variables
$\{1\}\cup\{2,3,4,5,6\}$	$i_1 \leq j_1 \leq j_2 \leq j_3 \leq j_4 \leq j_5$	$\{j_1,j_2,j_3,j_4,j_5\}$	$\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}$	$\{1,1,1,1,1\}$	i_1
$\{2\}\cup\{1,3,4,5,6\}$	$j_1 < i_1 \leq j_2 \leq j_3 \leq j_4 \leq j_5$	$\{i_1,j_2,j_3,j_4,j_5\}$	$\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}$	$\{1,1,1,1,1\}$	j_1
$\{3\}\cup\{1,2,4,5,6\}$	$j_1 \leq j_2 < i_1 \leq j_3 \leq j_4 \leq j_5$	$\{i_1,i_1,j_3,j_4,j_5\}$	$\mathbb{R}^2\oplus\mathbb{R}\oplus\mathbb{R}\oplus\mathbb{R}$	$\{2,1,1,1\}$	j_1,j_2
$\{4\}\cup\{1,2,3,5,6\}$	$j_1 \leq j_2 \leq j_3 < i_1 \leq j_4 \leq j_5$	$\{i_1,i_1,i_1,j_4,j_5\}$	$\mathbb{R}^3\oplus\mathbb{R}\oplus\mathbb{R}$	$\{3,1,1\}$	j_1,j_2,j_3
$\{5\}\cup\{1,2,3,4,6\}$	$j_1 \leq j_2 \leq j_3 \leq j_4 < i_1 \leq j_5$	$\{i_1,i_1,i_1,i_1,j_5\}$	$\mathbb{R}^4\oplus\mathbb{R}$	$\{4,1\}$	j_1,j_2,j_3,j_5
$\{6\}\cup\{1,2,3,4,5\}$	$j_1 \leq j_2 \leq j_3 \leq j_4 \leq j_5 < i_1$	$\{i_1,i_1,i_1,i_1,i_1\}$	$\mathbb{R}^5$	$\{5\}$	j_1,j_2,j_3,j_4,j_5

TABLE 3.

8. L^p-estimates for flag convolutions

In this section we establish the boundedness in $L^p(G)$ for $1 < p < \infty$ of the operator $f \to \mathcal{K} * f$ given by convolution on G with a flag kernel. To simplify the notation, we limit ourselves to the special situation where the exponents of the dilations $d_1, d_2, \ldots, d_n$ in equation (2.1) are positive integers. The results proved below will go over to the more general context with essentially no change in the proofs. We will also find it convenient to consider a continuous parameter s_k for the dilation of the x_k variable, in place of the dyadic version 2^{i_k} appearing in the previous sections. Again, the various results above stated for the dyadic dilations have simple modifications valid for their continuous analogues.

8.1. Maximal functions

As usual, G is a homogeneous nilpotent Lie group that we identify with $\mathbb{R}^N$ as in Section 6.6. We also let

$$G_k = \big\{\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^N \,\big|\, \mathbf{x}_1 = \cdots = \mathbf{x}_{k-1} = 0\big\}$$
$$= \big\{(\mathbf{0}, \ldots, \mathbf{0}, \mathbf{x}_k, \ldots, \mathbf{x}_n) \in \mathbb{R}^N \,\big|\, \mathbf{x}_j \in \mathbb{R}^{a_j}, \, k \leq j \leq n\big\}.$$

We can identify G_k with $\mathbb{R}^{a_k} \oplus \cdots \oplus \mathbb{R}^{a_n}$, and it follows from the formula (6.24) for group multiplication that G_k is a subgroup of G. We let $m(E)$ denote the Lebesgue measure of a set $E \subseteq G = G_1$, and $m_k(E)$ denote the Lebesgue measure on G_k of a subset $E \subseteq G_k$. For $\mathbf{s} = (s_k, \ldots, s_n)$, let

$$R_\mathbf{s} = R_\mathbf{s}^{(k)} = \{(\mathbf{x}_k, \ldots, \mathbf{x}_n) \in G_k : |\mathbf{x}_k| \leq s_k^k, \ldots, |\mathbf{x}_n| \leq s_n^n\}.$$

We say that the size of the rectangle $R_\mathbf{s}$ is *acceptable* if $s_k \leq s_{k+1} \leq \cdots \leq s_n$.

Definition 8.1. The maximal function M, defined on $G = G_1$, is given by

$$M(f)(\mathbf{x}) = \sup \frac{1}{m(R_{\mathbf{s}})} \int_{R_{\mathbf{s}}} |f(\mathbf{x} \cdot \mathbf{y}^{-1})| d\mathbf{y}$$

where the supremum is taken over all acceptable rectangles $R_{\mathbf{s}} = R_{\mathbf{s}}^{(1)} \subseteq G = G_1$.

Theorem 8.2.

(a) *M is a bounded map of $L^p(G)$ to itself, for $1 < p < \infty$.*

(b) *For $1 < p < \infty$ there are constants A_p so that if $\{f_j\}$ are scalar-valued functions on G then*

$$\left\|\left(\sum_j M(f_j)^2\right)^{1/2}\right\|_{L^p(G)} \leq A_p \left\|\left(\sum_j |f_j|^2\right)^{1/2}\right\|_{L^p(G)}.$$

To prove this theorem, we consider the standard maximal function M_k on the subgroup G_k defined by

$$M_k(f)(\mathbf{x}) = \sup_{\rho > 0} \frac{1}{m(B_\rho)} \int_{B_\rho} |f(\mathbf{x}\mathbf{y}^{-1})| \, d\mathbf{y}$$

where $B_\rho = B_\rho^{(k)}$ is the automorphic one-parameter ball given by

$$B_\rho^{(k)} = \left\{(\mathbf{x}_k, \ldots, \mathbf{x}_n) \in G_k : |\mathbf{x}_k| \leq \rho^k, |\mathbf{x}_{k+1}| \leq \rho^{k+1}, \ldots, |\mathbf{x}_n| \leq \rho^n\right\}.$$

Let $\widetilde{M}_k$ be the maximal function in G obtained by lifting M_k in G_k to G. (Facts about lifting are reviewed in the Appendix, Section 9). The key lemma is

Lemma 8.3. *There is a constant C so that*

$$M \leq C \, \widetilde{M}_n \circ \widetilde{M}_{n-1} \circ \cdots \circ \widetilde{M}_1 \, .$$

Proof. Let $\mathbf{s} = (s_k, s_{k+1}, \ldots, s_n)$ and $\bar{\mathbf{s}} = (s_{k+1}, \ldots, s_n)$. Let

$\chi_{R_{\mathbf{s}}^{(k)}}$ = the characteristic function of the rectangle $R_{\mathbf{s}}^{(k)}$ in the subgroup G_k,

$\chi_{R_{\bar{\mathbf{s}}}^{(k+1)}}$ = the characteristic function of $R_{\bar{\mathbf{s}}}^{(k+1)}$ in the subgroup G_{k+1},

$\chi_{B_{s_k}^{(k)}}$ = the characteristic function of the ball $B_{s_k}^{(k)}$ in G_k.

Let $\eta_{R_{\mathbf{s}}^{(k)}}, \eta_{R_{\bar{\mathbf{s}}}^{(k+1)}}, \eta_{B_{s_k}^{(k)}}$ be the normalized versions of these functions, so that, for example, $\eta_{R_{\mathbf{s}}^{(k)}} = m_k(R_{\mathbf{s}}^{(k)})^{-1} \chi_{R_{\mathbf{s}}^{(k)}}$, with a similar definition for $\eta_{B_{s_k}^{(k)}}$. The first observation to make is that if $s_k \leq s_{k+1} \leq \cdots \leq s_n$, there is an estimate $\eta_{R_{\mathbf{s}}^{(k)}} \lesssim \eta_{B_{s_k}^{(k)}} * \eta_{R_{\bar{\mathbf{s}}}^{(k+1)}}$, in the sense there are constants c, C, so that

$$(8.1) \qquad\qquad \eta_{R_{c\mathbf{s}}^{(k)}} \leq C \eta_{B_{s_k}^{(k)}} * \left(\delta_{\mathbf{x}_k} \otimes \eta_{R_{\bar{\mathbf{s}}}^{(k+1)}}\right)$$

where the convolution is now on the group G_k, and $\delta_{\mathbf{x}_k}$ denotes the delta function of the $\mathbf{x}_k$ variables.[10]

---

[10] Inequality (8.1) is essentially contained in Subsection 6.7.

In fact,

$$\chi_{B_{s_k}^{(k)}} * (\delta_{\mathbf{x}_k} \otimes \chi_{R_{\bar{\mathbf{s}}}^{(k+1)}}) = \int_{G_k} \chi_{B_{s_k}^{(k)}}(\mathbf{x} \cdot \mathbf{y}^{-1}) \chi_{R_{\bar{\mathbf{s}}}^{(k+1)}}(\mathbf{y}) \, d\mathbf{y}.$$

We introduce a new coordinate system in G_k, so that if $\mathbf{x} \in G_k$, then

$$\mathbf{x} = (\mathbf{x}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_n) = (\mathbf{x}_k, 0, \ldots, 0) \cdot (0, \mathbf{x}'_{k+1}, \ldots, \mathbf{x}'_n) = (\mathbf{x}_k) \cdot \mathbf{x}',$$

with $\mathbf{x}' \in G_{k+1}$. In this new coordinate system the integral can be written as

$$(8.2) \qquad \int_{G_{k+1}} \chi_{B_{s_k}^{(k)}}(\mathbf{x}_k \cdot \mathbf{y}') \chi_{R_{\bar{\mathbf{s}}}^{(k+1)}}(\mathbf{y}'^{-1} \cdot \mathbf{x}') \, d\mathbf{y}'.$$

Now if $\mathbf{x} \in R_{c\mathbf{s}}^{(k)}$, and $c > 0$ is small, then $|\mathbf{x}_k| \leq c^k s_k^k$ and $\mathbf{x}' \in R_{c'\bar{\mathbf{s}}}^{(k+1)}$, with c' small with c. (This is because $\mathbf{x}'_j = q_j(\mathbf{x})$, with q_j homogeneous polynomials of degree j, $k + 1 \leq j \leq n$.) So if $\mathbf{y}' \in B_{cs_{k+1}}^{(k+1)}$ then $\mathbf{y}'^{-1} \cdot \mathbf{x}' \in R_{\bar{\mathbf{s}}}^{(k+1)}$. Thus for $\mathbf{x} \in R_{c\mathbf{s}}^{(k)}$, the integrand above is 1, whenever $\mathbf{y}' \in B_{cs_{k+1}}^{(k+1)}$. The result is that the last integral exceeds $m_{k+1}(B_{cs_{k+1}}^{(k+1)})$ for $\mathbf{x} \in R_{c\mathbf{s}}^{(k)}$. Dividing through by the normalizing factors and observing that

$$\frac{1}{m_k(B_{s_k}^{(k)})} \cdot \frac{1}{m_{k+1}(R_{\bar{\mathbf{s}}}^{(k+1)})} \cdot m_{k+1}(B_{cs_{k+1}}^{(k+1)}) = c'' \frac{1}{m_k(R_{\mathbf{s}}^{(k)})}$$

proves the claim (8.1). Proceeding this way by downward induction, starting with the trivial case $k = n$, gives

$$(8.3) \qquad \eta_{R_{c\mathbf{s}}^{(1)}} \leq C \eta_{B_{s_1}^{(1)}} * (\delta_{\mathbf{x}_1} \otimes \eta_{B_{s_2}^{(2)}}) * \cdots * (\delta_{\mathbf{x}_1 \cdots \mathbf{x}_{n-1}} \otimes \eta_{B_{s_n}^{(n)}})$$

whenever $s_1 \leq s_2 \leq \cdots \leq s_n$. The inequality (8.3) then implies Lemma 8.3. $\square$

We now turn to the proof of Theorem 8.2. For each k, the maximal functions M_k satisfy the usual weak-type and L^p estimates on $L^p(G_k)$ (because the balls $B_{s_k}^{(k)}$ satisfy the required properties for the Vitali covering argument). Moreover, the vector-valued version

$$(8.4) \qquad \left\| \left(\sum_j (M_k(f_j))^2 \right)^{1/2} \right\|_{L^p(G_k)} \leq \mathcal{A}_p \left\| \left(\sum |f_j|^2 \right)^{1/2} \right\|_{L^p(G_k)}$$

also holds. This can be shown by following the main steps in the case of $\mathbb{R}^n$ (see, e.g., Chapter 2 of [11]). In fact, one proves first a weak-type inequality for the vector-valued case, using a Calderón–Zygmund decomposition, which establishes (8.4) for $1 < p \leq 2$. An additional argument is needed for $p > 2$, and is based on the fact that

$$\int_{G_k} M_k(f)^2(\mathbf{x}) \, \omega(\mathbf{x}) \, d\mathbf{x} \leq A \int_{G_k} |f(\mathbf{x})|^2 \, (M_k\omega)(\mathbf{x}) \, d\mathbf{x}$$

for all positive functions ω. Next a lifting argument (see the Appendix) allows one to lift (8.4) on G_k to G to get

$$(8.4') \qquad \left\| \left(\sum_j |\tilde{M}_k(f_j)|^2 \right)^{1/2} \right\|_{L^p(G)} \leq A_p \left\| \left(\sum_j |f_j|^2 \right)^{1/2} \right\|_{L^p(G)}.$$

As a result we obtain a similar inequality for $\tilde{M}_n \circ \tilde{M}_{n-1} \circ \cdots \circ \tilde{M}_1$, and an application of Lemma 8.3 then proves Theorem 8.2.

Our actual application of the estimate in equation $(8.4')$ is contained in the following:

Corollary 8.4. *Suppose $F_{\mathbf{t}}(\mathbf{x})$ is a measurable function of $(\mathbf{t}, \mathbf{x}) \in (\mathbb{R}^+)^n \times \mathbb{R}^N$. Then*

$$\left\| \left(\int_{(\mathbb{R}^+)^n} (M(F_{\mathbf{t}}))(\mathbf{x}))^2 d\mathbf{t} \right)^{1/2} \right\|_{L^p(\mathbb{R}^M)} \leq A_p \left\| \left(\int_{(\mathbb{R}^+)^n} |F_{\mathbf{t}}(\mathbf{x})|^2 d\mathbf{t} \right)^{1/2} \right\|_{L^p(\mathbb{R}^N)}.$$

Proof. Assume first that $F_{\mathbf{t}}(\mathbf{x})$ is jointly continuous and has compact support. For each $\epsilon > 0$, apply the conclusion $(8.4')$ to the case where $\{f_j(\mathbf{x})\}$ are an enumeration of the $\epsilon^{n/2} F_{\epsilon i_1, \epsilon i_2, \ldots, \epsilon i_n}(\mathbf{x})$, for $(i_1, i_2, \ldots, i_n)$ ranging over $(\mathbb{Z}^+)^n$, and then let $\epsilon \to 0$, obtaining the desired result in this case. For the general $F_{\mathbf{t}}$, assuming that $\|(\int_{(\mathbb{R}^+)^n} |F_{\mathbf{t}}(\mathbf{x})|^2 d\mathbf{t})^{1/2}\|_{L^p(G)}$ is finite, find a sequence $F_{\mathbf{t}}^{(n)}(\mathbf{x})$ of continuous functions of compact support, with $F_{\mathbf{t}}^{(n)}(\mathbf{x}) \to F_{\mathbf{t}}(\mathbf{x})$ almost everywhere, so that

$$\left\| \left(\int_{(\mathbb{R}^+)^n} |F_{\mathbf{t}}^{(n)}|^2 d\mathbf{t} \right)^{1/2} \right\|_{L^p} \to \left\| \left(\int_{(\mathbb{R}^+)^n} |F_{\mathbf{t}}|^2 d\mathbf{t} \right)^{1/2} \right\|_{L^p}$$

and apply the previous case, via Fatou's lemma. □

It will also be useful to observe that effectively the estimate (8.3) can be reversed in the following way.

Lemma 8.5. *We have*

$$(8.5) \qquad \eta_{B_{s_1}^{(1)}} * (\delta_{\mathbf{x}_1} \otimes \eta_{B_{s_2}^{(2)}}) * \cdots * (\delta_{\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}} \otimes \eta_{B_{s_n}^{(n)}}) \leq C \eta_{R_{c\mathbf{s}^*}^{(1)}}$$

for an appropriate $C > 0$. Here $\mathbf{s}^* = (s_1^*, \ldots, s_n^*)$, with $s_k^* = \max\{s_j, 1 \leq j \leq k\}$. Note that we do not require that $s_1 \leq s_2 \leq \cdots \leq s_n$.

The proof is based on the observation that $\eta_{B_{s_k}^{(k)}} * (\delta_{\mathbf{x}_k} \otimes \eta_{R_{\tilde{\mathbf{s}}}^{(k+1)}}) \leq c \eta_{R_{c\tilde{\mathbf{s}}}^{(k)}}$, where $\tilde{\mathbf{s}} = (\tilde{s}_k, \ldots, \tilde{s}_n)$ and

$$\tilde{s}_j = \begin{cases} \max\{s_k, s_j\} & \text{if } j > k, \\ s_k & \text{if } j = k. \end{cases}$$

In fact if $\mathbf{x} \notin R_{C\tilde{\mathbf{s}}}^{(k)}$ (for some large C) then either $|\mathbf{x}_k| \geq C' s_k^k$ or $\mathbf{x}' \notin R_{c\tilde{\mathbf{s}}}^{(k+1)}$. Looking back at the integral in equation (8.2) we see that the integral vanishes, because $\chi_{B_{s_k}^{(k)}}(\mathbf{x}_k \cdot \mathbf{y}') = 0$ in the first case, or $\chi_{R_{\tilde{\mathbf{s}}}^{(k+1)}}(\mathbf{y}'^{-1} \cdot \mathbf{x}') = 0$ in the second case. Moreover as above, this integral is majorized by $C m(B_{cs_{k+1}}^{(k+1)})$. Altogether then, we have $\eta_{B_{s_k}^{(k)}} * (\delta_{\mathbf{x}_k} \otimes \eta_{R_{\tilde{\mathbf{s}}}^{(k+1)}}) \leq c \eta_{R_{\tilde{\mathbf{s}}}^{(k)}}$, and an induction proves Lemma 8.5.

Now let $A_\mathbf{s}$ denote the function appearing on the left-side of (8.5). Then as a consequence we have

$$(8.6) \qquad |(f * A_\mathbf{s})(\mathbf{x})| \leq C\, M(f)(\mathbf{x}),$$

for all $\mathbf{s} = (s_1, \ldots, s_n)$, not necessarily in increasing order. Similarly if $A_\mathbf{s}^* = (\delta_{\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}} \otimes \eta_{B_{s_n}^{(n)}}) * \cdots * (\eta_{B_{s_1}^{(1)}})$ we also have $A_\mathbf{s}^*(\mathbf{x}) \leq c\, \eta_{R_{\mathbf{s}^*}^{(1)}}$. This follows from (8.5) if we observe that $A_\mathbf{s}^*(\mathbf{x}) = A_\mathbf{s}(\mathbf{x}^{-1})$. As a result, in analogy to (8.6), we have for all $\mathbf{s}$

$$(8.6') \qquad |(f * A_\mathbf{s}^*)(\mathbf{x})| \leq c\, M(f)(\mathbf{x}).$$

Indeed, one has $R_{\mathbf{s}^*}^{-1} = R_{\mathbf{s}^*}$ if, in defining $R_\mathbf{s}$, a coordinate system is used where the inverse of $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is given by $\mathbf{x}^{-1} = (-\mathbf{x}_1, \ldots, -\mathbf{x}_n)$. Alternatively, if we use canonical coordinates of the second kind, as above, then one has $R_{c_1 \mathbf{s}^*} \subset R_{\mathbf{s}^*}^{-1} \subset R_{c_2 \mathbf{s}^*}$, for two appropriate constants c_1 and c_2; this also leads to (8.6').

8.2. Comparisons

The basic comparison function is

$$\Gamma_\mathbf{t}(\mathbf{x}) = t_1 \cdot t_2 \cdots t_n \cdot \prod_{k=1}^n (t_1 + t_2 + \cdots + t_k + N_1(\mathbf{x}) + \cdots + N_k(\mathbf{x}))^{-Q_k - 1}$$

for $\mathbf{t} = (t_1, \ldots, t_n)$, $t_j > 0$. Recall that $N_k(\mathbf{x}) = |\mathbf{x}_k|^{1/k}$ and $Q_k = k a_k$ with a_k the dimension of the x_k space.

Theorem 8.6.

$$(8.7) \qquad \sup_\mathbf{t} \left|(f * \Gamma_\mathbf{t})(\mathbf{x})\right| \leq C\, M(f)(\mathbf{x}),$$

where the supreme is taken over all $\mathbf{t}$, with $t_j > 0$.

Proof. Note that it suffices to restrict attention to $\mathbf{t}$'s that are of acceptable size. Indeed, let $s_j = t_1 + t_2 \cdots + t_j$, $1 \leq j \leq k$. Then $s_j \leq s_{j+1}$ but $k(t_1 + \cdots + t_k) \geq s_1 + s_2 + \cdots + s_k$. Hence,

$$(t_1 + \cdots + t_k + N_1 + \cdots + N_k)^{-Q_k - 1} \leq c_k (s_1 + s_2 \cdots + s_k + N_1 + \cdots + N_k)^{-Q_k - 1},$$

with $c_k = k^{Q_k + 1}$. Therefore $\Gamma_\mathbf{t}(\mathbf{x}) \leq c\, \Gamma_\mathbf{s}(\mathbf{x})$, which shows that it suffices to consider $\mathbf{t}$'s that are increasing.

We next fix $\mathbf{t} = (t_1, \ldots, t_n)$ and decompose the space $G = \mathbb{R}^N$ into a preliminary dyadic partition as follows. For each $J = (j_1, \ldots, j_n) \in \mathbb{Z}_+^n$ we let

$$R_J = \left\{ \mathbf{x} \in \mathbb{R}^N \,\Big|\, 2^{j_k - 1} < \frac{N_1(\mathbf{x}) + \cdots + N_k(\mathbf{x})}{t_1 + t_2 + \cdots + t_k} \leq 2^{j_k}, \text{ for } k = 1, 2, \ldots, n \right\},$$

with the understanding that if $j_k = 0$ the inequality should be taken to be

$$\frac{N_1(\mathbf{x}) + \cdots + N_k(\mathbf{x})}{t_1 + t_2 + \cdots + t_k} \leq 1.$$

Notice that $\bigcup_{J\in\mathbb{Z}_+^n} R_J = G$ gives a partition of the space G. However, in general each R_J is not comparable to an acceptable rectangle. We remedy this as follows. For a suitable constant c, define $s^J = (s_1^J, \ldots, s_n^J)$ by

(8.8) $s_1^J = ct_1 2^{j_1}, \quad s_2^J = c^2 t_2 2^{j_2}, \quad \ldots, \quad s_n^J = c^n t_n 2^{j_n}.$

Now if R_J is non-empty, then since $t_1 \le t_2 \le \cdots \le t_n$, we have

$$N_1(\mathbf{x}) \approx t_1 2^{J_1}, \quad N_1(\mathbf{x}) + N_2(\mathbf{x}) \approx t_2 2^{j_2}, \quad \ldots, \quad N_1(\mathbf{x}) + N_2(\mathbf{x}) + \cdots + N_n(\mathbf{x}) \approx t_n 2^{j_n}.$$

As a result, for sufficiently large c, it follows that $s_1^J \le s_2^J \le \cdots \le s_n^J$.

Now define $R_J^* = \{\mathbf{x} : N_k(\mathbf{x}) \le s_k^J, k = 1, \ldots, n\}$ for those J where R_J is not empty. Then clearly $R_J \subset R_J^*$ and each R_J^* is a rectangle of acceptable size (in fact, essentially the smallest rectangle of acceptable size containing R_J). However, for $f \ge 0$,

$$\int_G f(\mathbf{x} \cdot \mathbf{y}^{-1}) \, \Gamma_\mathbf{t}(\mathbf{y}) \, d\mathbf{y} = \sum_{J\in\mathbb{Z}_+^n} \int_{R_J} f(\mathbf{x}\mathbf{y}^{-1}) \, \Gamma_\mathbf{t}(\mathbf{y}) \, d\mathbf{y}.$$

Recall that by (8.8), $t_k \approx s_k^J 2^{-j_k}$, and on R_J we have $N_1 + \cdots + N_k \approx (t_1 + t_2 + \cdots + t_k) 2^{j_k} \approx s_k^J$. Thus, on R_J we have $\Gamma_\mathbf{t}(\mathbf{y}) \lesssim \prod_{k=1}^n (s_k^J)^{-Q_k} \cdot (2^{-Q_k j_k})$. So

$$\int_G f(\mathbf{x}\mathbf{y}^{-1})\Gamma_\mathbf{t}(\mathbf{y})d\mathbf{y} \lesssim \sum_J \prod_{k=1}^n 2^{-Q_k j_k}(s_k^J)^{-Q_k} \cdot \int_{R_J^*} f(\mathbf{x}\mathbf{y}^{-1})d\mathbf{y}.$$

However, each R_J^* is a rectangle of acceptable size and $\prod_{k=1}^n (s_k^J)^{Q_k} = c\,m(R_J^*)$. Thus by the definition of M, the last sum is majorized by

$$c \sum_J \prod_{k=1}^n 2^{-Q_k j_k} M(f) = c' M(f),$$

and (8.7) is proved.   □

8.3. Truncated kernels

Recall that we defined truncated kernels and improved truncated kernels in Definition 6.20 (Section 6.9). Suppose that $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^N)$ with support in the unit ball. For $b > 0$ write

$$\psi_b(\mathbf{x}) = b^{-Q_1 - Q_2 - \cdots - Q_n} \, \psi(b^{-1}\mathbf{x}_1, b^{-2}\mathbf{x}_2, \ldots, b^{-n}\mathbf{x}_n),$$

the automorphically dilated ψ. We also say that ψ_b has *width* b.

Theorem 8.7. *Suppose $\mathcal{K}$ is a truncated flag kernel of width a, and ψ_b is as above of width b. Then*

(1) *$\mathcal{K} * \psi_b$ and $\psi_b * \mathcal{K}$ are truncated kernels of width $a + b$.*

(2) *If in addition $\int_G \psi(\mathbf{x}) \, d\mathbf{x} = 0$, then $K * \psi_b$ and $\psi_b * K$ are improved truncated kernels of width $a+b$. Moreover, then $K * \psi_b$ and $\psi_b * K$ are actually improved truncated kernels of width $a + b$, multiplied by the further factor $\frac{b}{a+b}$.*

Note that the statements of the hypotheses and conclusions have an automorphic-dilation invariance, so in proving Theorem 8.7, it suffices to consider two cases: $b = 1$, $a \leq 1$; and $a = 1$, $b \leq 1$. In the first case we use Proposition 6.21, since any truncated kernel is actually an un-truncated kernel. Thus we get that $K * \psi_1$ and $\psi_1 * K$ have width 1, which is essentially the same as having width $1 + a$, since $a \leq 1$.

The second case would be easy if every truncated kernel of width 1 were of the form $K * \psi_1$. Its proof is a little more involved and requires the following lemma.

Lemma 8.8. *Given any M, there exist η_0 and η_1 both of class $C^{(M)}$, supported in the unit ball, and a (non-communicative) polynomial $P(X_1, \ldots, X_N) = P(X)$ in the right-invariant vector fields of G, so that*

$$(8.9) \qquad\qquad P(X)\eta_0 = \delta_0 + \eta_1.$$

with δ_0 the Dirac delta at the origin.

Proof. Consider the elliptic operator of order $2r$, $P(X) = \left(\sum_{j=1}^{N} X_j^2\right)^r$, with r a positive integer. Then by the standard theory of pseudo-differential operators there is a locally integrable function F which is C^∞ away from the origin, so that $P(X)F = \delta_0 + \eta'$, with η' a C^∞ function. Moreover, F satisfies the estimate $\left| \left(\frac{\partial}{\partial x}\right)^\alpha F(\mathbf{x}) \right| \leq A_\alpha$, whenever $|\mathbf{x}| \leq 1$ and $2r > N + |\alpha|$ (these estimates also follow from Theorem 1 in [8]). Thus we only need to take $2r > M + N$ and set $\eta_0 = \mu \cdot F$, where μ is a C^∞ function supported in the unit ball, and $\mu(x) = 1$ in the ball of radius $1/2$. Then since η_0 is supported in the unit ball, so is $\eta_1 = P(X)\eta_0 - \delta_0$; and since F is C^∞ away from the origin it follows that η_1 is in fact C^∞ everywhere. This completes the proof of Lemma 8.8. $\qquad\square$

We now return to the proof of Theorem 8.7. We consider $K * \psi_b$ when K has width 1, and $b \leq 1$. Now by the lemma $K * \psi_b = K * \delta_0 * \psi_b = K * P(X) * \eta_0 * \psi_b + K * \eta_1 * \psi_b$, since $P(X)$ is a right-invariant differential operator. Now since K has width 1, $K * P(X)$ is also a truncated kernel of width one, and in particular an un-truncated kernel. However, $\eta_0 * \psi_b$ has width $1 + b$, which is essentially one. Also, it is of class $C^{(M)}$ (uniformly in b), since η_0 is of class $C^{(M)}$. Thus $K * P(X) * \eta_0 * \psi_b$ satisfies the differential inequalities for a truncated kernel of width one for all orders $\leq m$. However, the term $K * \eta_1 * \psi_b$ clearly does the same, for all orders. Notice we can make m as large as we wish by making M sufficiently large. (See Remark 6.22 on page 690.) A similar argument works for $\psi_b * K$ and thus part (1) of Theorem 8.7 is proved.

Part (2) is proved in the same way, using conclusion (2) of Proposition 6.21. The further improvement given by the factor $b(a + b)^{-1}$ comes about as follows. As before, we may take $a = 1$, and $b \leq 1$. Since $\int \psi(\mathbf{x}) \, d\mathbf{x} = 0$, both $\eta_0 * \psi_b$ and $\eta_1 * \psi_b$ give an improvement of b. In fact, since $\int \psi(\mathbf{x}) \, d\mathbf{x} = 0$, it follows from Lemma 5.1 and Proposition 4.5 that we can write

$$(8.10) \qquad\qquad \psi_b = \sum b^k X_k(\psi_b^{(k)})$$

for suitable C^∞ functions $\psi^{(k)}$ supported in the unit ball, with $\{X_k\}$ ranging over right-invariant vector fields of degree k , $k \geq 1$. Thus $\eta_0 * \psi_b = \sum b^k (\eta_0 * X_k) * \psi_b^{(k)}$ and this gives a gain b, $b \leq 1$. Similarly for the term $\eta_1 * \psi_b$.

8.4. Key estimates: kernels

Suppose $\varphi^{(k)} \in \mathcal{C}_0^\infty$ is supported on the unit ball of the group G_k, with

$$\int_{G_k} \varphi^{(k)}(\mathbf{x})\, d\mathbf{x} = 0.$$

We set $\varphi_t^{(k)}(\mathbf{x}) = t^{-Q_k - \cdots - Q_n} \varphi(\delta_{t^{-1}}(\mathbf{x}))$, with $Q_k = k a_k$ and let $\tilde{\varphi}^{(k)}$ to be the corresponding distributions lifted to the full group G; *i.e.*, $\tilde{\varphi}_t^{(k)} = \delta_{\mathbf{x}_1, \mathbf{x}_2 \cdots \mathbf{x}_{k-1}} \otimes \varphi_t^{(k)}$. We let $\Phi_{\mathbf{t}} = \tilde{\varphi}_{t_1}^{(1)} * \tilde{\varphi}_{t_2}^{(2)} * \cdots * \tilde{\varphi}_{t_n}^{(n)}$ for $\mathbf{t} = (t_1, \ldots, t_n)$, and write

$$\Phi_{\mathbf{t}}^* = \tilde{\varphi}_{t_n}^{(n)} * \tilde{\varphi}_{t_{n-1}}^{(n-1)} * \cdots * \tilde{\varphi}_{t_1}^{(1)}.$$

Recall the comparison function $\Gamma_{\mathbf{t}}$ discussed in Section 8.2. Note that here we will allow the functions φ and Φ to take their values in finite-dimensional vector spaces.

Theorem 8.9. *Suppose $\mathcal{K}$ is a flag kernel. Then*

(1) $|K * \Phi_{\mathbf{t}}(\mathbf{x})|$ *and* $|\Phi_{\mathbf{t}}^* * K(\mathbf{x})|$ *are both majorized by* $c\, \Gamma_{\mathbf{t}}(\mathbf{x})$ *for all* $\mathbf{t}$.

(2) *If* X_k^R *is any right-invariant vector field of degree k, then*

$$|X_k^R(K * \Phi_{\mathbf{t}})| \leq c\,(t_1 + \cdots + t_k)^{-k}\, \Gamma_{\mathbf{t}}(\mathbf{x}).$$

(3) *If* X_k^L *is any left-invariant vector field of degree k, then*

$$|X_k^L(K * \Phi_{\mathbf{t}})| \leq c\,(t_1 + \cdots + t_k)^{-k}\, \Gamma_{\mathbf{t}}(\mathbf{x}).$$

For the proof we need to do our calculations in a particular coordinate system, already used in the proof Lemma 8.3. Here we represent a point $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in G$ via adapted canonical coordinates of the second kind; *i.e.*, we take $\mathbf{x} = \exp(\mathbf{x}_1' \cdot X_1) \exp(\mathbf{x}_2' \cdot X_2) \cdots \exp(\mathbf{x}_n' \cdot X_n)$, where $\mathbf{x}' = (x_{k,1}', \ldots, x_{k,a_k}')$, $\{X_{k,1}, \ldots, X_{k,a_k}\}$ is a basis of the subspace of vector fields of degree k, and $\mathbf{x}_k' \cdot X_k = \sum_j x_{k,j}' X_{k,j}$. The passage from the initial $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ coordinates to the $(\mathbf{x}_1', \ldots, \mathbf{x}_n')$ coordinates is of the form treated in Section 6.5, so that the basic comparison function $\Gamma_{\mathbf{t}}$ is essentially unchanged when passing from $\mathbf{x}$ to $\mathbf{x}'$. We therefore freely use instead the new coordinate system $\mathbf{x}'$, and now relabel $\mathbf{x}'$ by $\mathbf{x}$. The advantage of this coordinate system is that, firstly $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_k, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_n) \in G_k$ if and only if $\mathbf{y}_1 = 0, \ldots, \mathbf{y}_{k-1} = 0$, but more importantly, if $\mathbf{x} \in G$, and $\mathbf{y} \in G_k$, then $\mathbf{x} \cdot \mathbf{y} = (\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, \bar{\mathbf{x}}_k, \ldots, \bar{\mathbf{x}}_n)$, with $\bar{\mathbf{x}}_\ell$ (for $\ell \geq k$), depending only on $\mathbf{x}_k, \ldots, \mathbf{x}_n$, and $\mathbf{y}_k, \ldots, \mathbf{y}_n$, and not on $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$.

Now set $K^{(k)} = K * \tilde{\varphi}_{t_1}^{(1)} * \cdots * \tilde{\varphi}_{t_k}^{(k)}$. Consider first $K^{(1)} = K * \varphi_{t_1}^{(1)}$. According to Proposition 6.21, part (2), since $\int_G \varphi^{(1)}(\mathbf{x})\, d\mathbf{x} = 0$, then for each $\mathbf{x}_1$, the kernel

$K^{(1)}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, as a function of $(\mathbf{x}_2, \ldots, \mathbf{x}_n)$ on G_2, is a truncated kernel of width $t_1 + N_1(\mathbf{x}_1)$, multiplied by the "constant" factor $t_1 \left[t_1 + N_1(\mathbf{x}_1)\right]^{-Q_1 - 1}$. Also we have a similar conclusion for $\partial_{x_1}^{\beta_1} K^{(1)}(\mathbf{x})$, except now the improving factor is $t_1 \left[t_1 + N_1(\mathbf{x}_1)\right]^{-Q_1 - 1 - \beta_1}$.

Consider next the inductive hypothesis: for a given k,

(a) for each $\mathbf{x}_1, \ldots, \mathbf{x}_k$, $K^{(k)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$, thought of as a function of $(\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n)$ on G_{k+1}, is a truncated kernel of width $t_1 + \cdots + t_k + N_1(\mathbf{x}_1) + \cdots + N_k(\mathbf{x}_k)$, multiplied by the improving factor $\prod_{j=1}^k t_j \left[t_1 + \cdots + t_j + N_1 + \cdots + N_j\right]^{-Q_j - 1}$;

(b) for each $r \le k$, a similar statement holds for $\partial_{x_r}^{\beta_r} K^{(k)}(\mathbf{x})$, except that now the rth part of the improving factor is $t_r \left[t_1 + \cdots + t_r + N_1 + \cdots + N_r\right]^{-Q_r - 1 - r\beta_r}$.

Notice that if the inductive hypothesis holds for k, that is for $K^{(k)}$, then since $K^{(k+1)}(\mathbf{x}_1 \cdots \mathbf{x}_n) = K^{(k)} * \varphi_{t_{k+1}}^{(k+1)}$, where since the convolution is taken on the group G_{k+1}, it therefore does not involve the variables $\mathbf{x}_1, \ldots, \mathbf{x}_k$ because of the nature of our coordinate system. As a result, we get the conclusion for $k+1$, that is for $K^{(k+1)}$. To see this, we merely apply Proposition 6.21, part (2), for the case of the group G_{k+1}.

More precisely, we are convolving a truncated kernel of width $a = t_1 + \cdots + t_k + N_1 + \cdots + N_k$ (on G_{k+1}) with a function φ (which equals $\varphi_{t_{k+1}}^{(k+1)}$) of width $b = t_{k+1}$. The result is a truncated kernel of width $a + b = t_1 + \cdots + t_{k+1} + N_1 + \cdots + N_k$ on G_{k+1}, together with a further factor $\frac{b}{a+b}$. That is, for $K^{(k+1)}$ as a function of $\mathbf{x}_{k+2}, \ldots, \mathbf{x}_n$, we have a truncated kernel of width $t_1 + \cdots + t_{k+1} + N_1 + \cdots + N_{k+1}$, times a factor of the form $b(a+b)^{-1}$. So the full improvement is

$$\frac{b}{a+b} \cdot \frac{a+b}{(a+b+N_{k+1})^{Q_{k+1}+1}} = \frac{t_{k+1}}{(t_1 + \cdots + t_{k+1} + N_1 + \cdots + N_{k+1})^{Q_{k+1}+1}},$$

as was needed. The same kind of improvement holds for the estimates of $\partial_{x_l}^{\beta_\ell} K^{(k+1)}(\mathbf{x})$, for $\ell \le k + 1$.

Thus, the inductive hypothesis (now the conclusion of the induction) holds for $k = n$. As a result, it is clear that $|K * \Phi_\mathbf{t}(\mathbf{x})| \le c\, \Gamma_\mathbf{t}(\mathbf{x})$ and

$$|\partial_{\mathbf{x}_k}(K * \Phi_\mathbf{t}(\mathbf{x}))| \le c \left[t_1 + \cdots + t_k + N_1 + \cdots + N_k\right]^{-k} \Gamma_\mathbf{t}(\mathbf{x}),$$

for every k, $1 \le k \le n$. Since $X_k^R = \partial_{\mathbf{x}_k} + \sum_{\ell > k} h_\ell^k \, \partial_{\mathbf{x}_\ell}$ where h_ℓ^k is a homogeneous polynomial of degree $\ell - k$, it follows, in particular, that

$$X_k^R(K * \Phi_\mathbf{t})(\mathbf{x}) \le c \left[t_1 + \cdots + t_k\right]^{-k} \Gamma_\mathbf{t}(\mathbf{x}).$$

The results for $\Phi_\mathbf{t}^* * K$ and $X_k^L(\Phi_\mathbf{t}^* * K)$ follow in the same way, but require a canonical coordinate system in the reverse order. Alternatively we can deduce them from the previous case by using the inversion $\mathbf{x} \to \mathbf{x}^{-1}$.

8.5. Key estimates: operators

We define $P_{\mathbf{t}}(f) = f * \Phi_{\mathbf{t}}$, and $P_{\mathbf{t}}^* = f * \Phi_t^*$, with $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ and $t_j > 0$, $1 \leq j \leq n$, with $\Phi_{\mathbf{t}}$ and $\Phi_{\mathbf{t}}^*$ defined at the beginning of the previous section. We suppose $\mathcal{K}$ is a flag kernel and $Tf = f * K$ when f is a Schwartz function. We recall the maximal operator M and let $\mathcal{M} = M \circ M$, i.e., $\mathcal{M}(f) = M(M(f))$.

Theorem 8.10.

(a) $|P_{\mathbf{t}} T(f)(\mathbf{x})| \leq c\, M(f)(\mathbf{x})$, all $\mathbf{t}$.

(a′) Similarly, $|(T P_{\mathbf{t}}^*)(f)(\mathbf{x})| \leq c\, M(f)(\mathbf{x})$, all $\mathbf{t}$.

(b) $|P_{\mathbf{t}} T P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq \gamma(\mathbf{s}, \mathbf{t})\, \mathcal{M}(f)(\mathbf{x})$, where, for some $\delta > 0$,

$$\gamma(\mathbf{s}, \mathbf{t}) \leq c \left(\prod_{k=1}^{n} \min \left(\frac{s_k}{t_k}, \frac{t_k}{s_k} \right) \right)^{\delta}.$$

Note: the conclusion will be seen to hold for $\delta = 1/n^2$.

Proof. The function $P_{\mathbf{t}} T(f)$ is given by $(f * K) * \Phi_{\mathbf{t}} = f * (K * \Phi_{\mathbf{t}})$. Hence conclusion (a) is a direct consequence of Theorem 8.6 and Theorem 8.9, part (1). The same is true for conclusion (a′).

Turning to (b), we first fix k, and consider the situation when $t_k/s_k = \rho \geq 1$. With this ρ given, we next divide our consideration in two cases.

Case I: With σ a positive constant, to be specified below, $s_{j-1}/s_j > \rho^{\sigma}$ for at least one j, with $2 \leq j \leq k$.

Case II: $s_{j-1}/s_j \leq \rho^{\sigma}$ for all j, with $2 \leq j \leq k$.

To handle Case I we need the following observation, that will give us the needed gain. Recall the notation $\eta_{B_\rho^{(k)}} = m_k(B_\rho^{(k)})^{-1} \chi_{B_\rho^{(k)}}$ used above in Section 8.1.

Lemma 8.11. If $s_{j-1}/s_j \geq \rho^{\sigma}$, $\rho \geq 1$, then

$$\left| \tilde{\varphi}_{s_j}^{(j)} * \varphi_{s_{j-1}}^{(j-1)} \right| \leq c\, \rho^{-\sigma j} \left(\tilde{\eta}_{B_{s_j}^{(j)}}^{(j)} \right) * \left(\eta_{B_{s_{j-1}}^{(j-1)}}^{(j-1)} \right).$$

Proof. In analogy with equation (8.10) in Section 8.3, we have that on G_j

$$(8.11) \qquad \varphi_{s_j}^{(j)} = \sum_{r \geq j} (s_j)^r\, X_r^L(\psi_{s_j}^{(r)})$$

where X_r^L are left-invariant vector fields on G_j, with $\psi^{(r)}$ C^{∞} functions supported in the unit ball. Now

$$\left(\delta_{\mathbf{x}_{j-1}} \otimes X_r^L(\psi_{s_j}^{(r)}) \right) * \varphi_{s_{j-1}}^{(j-1)} = \left(\delta_{\mathbf{x}_{j-1}} \otimes \psi_{s_j}^{(r)} \right) * X_r^R\left(\varphi_{s_{j-1}}^{(j-1)} \right),$$

and $X_r^R\left(\varphi_{s_{j-1}}^{(j-1)} \right)$ is of the form $s_{j-1}^{-r}\, \psi_{s_{j-1}}'^{(r)}$. Combining these gives a sum

$$\sum_{r \geq j} (s_j/s_{j-1})^r\, \tilde{\psi}_{s_j}^{(r)} * \psi_{s_{j-1}}'^{(r)}$$

and since $s_j/s_{j-1} \leq \rho^{-\sigma}$, the lemma is proved. $\qquad\square$

Consider the operator $P_{\mathbf{s}}^*$ given by $P_{\mathbf{s}}^*(f) = f * \Phi_{\mathbf{s}}^*$, with $\Phi_{\mathbf{s}}^* = \tilde{\varphi}_{s_n}^{(n)} * \tilde{\varphi}_{s_{n-1}}^{(n-1)} * \cdots *$ $\tilde{\varphi}_{s_1}^{(1)}$. Because of Lemma 8.11 and since $j \geq 2$, we have $|\Phi_{\mathbf{s}}^*| \leq c\rho^{-2\sigma} \tilde{\eta}_{B_{s_n}^{(k)}} * \cdots *$ $\tilde{\eta}_{B_{s_1}^{(1)}} = c\,\rho^{-\sigma j} = c\rho^{-2\sigma} A_{\mathbf{s}}^*$, in the notation of Section 8.1. So by equation (8.6$'$) there, it follows that $|P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq c\rho^{-2\sigma} M(f)(\mathbf{x})$. Combining this with the first conclusion already proved that $|P_{\mathbf{t}} T(F)(\mathbf{x})| \leq cM(F)(\mathbf{x})$, with $F = P_{\mathbf{t}}^*(f)$, yields

$$(8.12) \qquad |P_{\mathbf{t}} T P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq c\rho^{-2\sigma} \mathcal{M}(f)(\mathbf{x})$$

since $\mathcal{M} = M \cdot M$.

We now turn to Case II. Here $s_{j-1}/s_j \leq \rho^\sigma$, for all j, $2 \leq j \leq k$. Thus $s_j \leq s_k \rho^{\sigma(k-j)}$ for $1 \leq j \leq k$. If we set $s_* = \sup_{1 \leq j \leq k} s_j$, then $s_* \leq s_k \rho^{\sigma(k-1)}$. Next we recall the following fact:

If X_ℓ^R is a right-invariant vector field of degree $\ell \geq j$ on G_j, then one can write

$$X_\ell^R(\psi_a) = \sum_{r \geq \ell} a^{r-\ell} X_r^L(\psi_a)\,,$$

where X_r^L are left-invariant vector fields of degree r. This follows by writing $X_\ell^R = \sum h_{\ell,r}(\mathbf{x}) X_r^L$, with $h_{\ell,r}(\mathbf{x})$ a homogeneous polynomial of degree $r - \ell$, and arguing as in Proposition 3.2.

With this in hand, consider $\tilde{\varphi}_{s_k}^{(k)} * \tilde{\varphi}_{s_{k-1}}^{(k-1)} * \cdots * \tilde{\varphi}_{s_1}^{(1)}$. First use that $\int_{G_k} \varphi^{(k)}(\mathbf{x}) d\mathbf{x} = 0$, which by the analogue of assertion (8.10) (for left-invariant vector fields on the group G_k) gives an expression involving the action of left-invariant vector fields (of degrees $\geq k$). Next pass from the left-invariant vector-fields acting on $\psi_{s_j}^{(r)}$ to the corresponding right-invariant vector fields acting on $\tilde{\varphi}_{s_{k-1}}^{k-1}$, via the rule

$$(8.13) \qquad (X_r^L \psi) * \varphi = \psi * X_r^R \varphi.$$

At this point utilize the remark on page 714 to pass to left-invariant vector fields, and then use the rule (8.13) above to pass to $\tilde{\varphi}_{s_{k-2}}^{k-2}$, etc.

Putting this all together leads quickly to the following conclusion: the convolution

$$\tilde{\varphi}_{s_k}^{(k)} * \tilde{\varphi}_{s_{k-1}}^{(k-1)} * \cdots * \tilde{\varphi}_{s_1}^{(1)}$$

is a finite sum of expression of the form

$$(8.14) \qquad (s_*)^r \tilde{\psi}_{s_k}^{(k)} * \tilde{\psi}_{s_{k-1}}^{(k-1)} * \cdots * X_r^L(\tilde{\psi}_{s_1}^{(1)})$$

for $r \geq k$. Here we use the fact that $s_* \geq s_j$, $j \leq k$. Now consider

$$(8.15) \qquad \tilde{\varphi}_{s_k}^{(k)} * \tilde{\varphi}_{s_{k-1}}^{(k-1)} * \cdots * \tilde{\varphi}_{s_1}^{(1)} * K * \Phi_t.$$

By applying the rule (8.3) we can pass the left-invariant vector field X_r^L as a right-invariant vector field acting on $K * \Phi_{\mathbf{t}}$. We keep in mind that $r \geq k$, and use Theorem 8.9, part (2). Therefore $\Phi_{\mathbf{s}}^* * K * \Phi_{\mathbf{t}} = \tilde{\varphi}_{s_n}^{(n)} * \tilde{\varphi}_{s_{k-1}}^{(k-1)} * \cdots * \tilde{\varphi}_{s_1}^{(1)} * K * \Phi_{\mathbf{t}}$ is majorized by a constant multiple of $s_*^r \cdot t_k^{-r} A_{\mathbf{s}}^* * \Gamma_{\mathbf{t}}$, where $A_{\mathbf{s}}^* = \tilde{\eta}_{B_{s_n}^{(n)}} * \tilde{\eta}_{B_{s_{n-1}}^{(n-1)}} * \cdots * \tilde{\eta}_{B_{s_1}^{(1)}}$.

In view of Theorem 8.6 and inequality (8.6′) in Section 8.1, we get that

$$(8.16) \qquad |P_{\mathbf{t}} T P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq c \sup_{r \geq k} (s_*^r t_k^{-r}) \, M(M(f))(\mathbf{x}).$$

We now pick $\sigma = 1/2k$. Since $s_* \leq s_k \rho^{\sigma(k-1)}$, and $s_k/t_k = \rho^{-1}$, $\rho \geq 1$, we get

$$|P_{\mathbf{t}} T P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq c \, \rho^{-1/2} \, M(M(f))(\mathbf{x}) \, .$$

Combining this with the previous case given by (8.12) yields

$$|P_{\mathbf{t}} T P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq c \Big(\frac{s_k}{t_k} \Big)^{1/n} \mathcal{M}(f)(\mathbf{x}), \quad \text{if } t_k/s_k \geq 1.$$

By a parallel argument the analogous result holds for $s_k/t_k \geq 1$. Hence

$$|P_{\mathbf{t}} T P_{\mathbf{s}}^*(f)(\mathbf{x})| \leq c \, \min \Big(\frac{s_k}{t_k}, \, \frac{t_k}{s_k} \Big)^{\frac{1}{n}} \mathcal{M}(f)(\mathbf{x}) \, .$$

Since this holds for all k, $1 \leq k \leq n$, we can take the geometric mean of these inequalities. The result is conclusion (b) of Theorem 8.10, with $\delta = 1/n^2$.   □

8.6. Square functions and L^p-boundedness

We will construct the square functions for G as products of the (one-parameter) square functions of the subgroups G_k, $1 \leq k \leq n$. Each G_k is a homogeneous group with family of dilations δ_r, and so there exists a finite-dimensional inner product space V_k and a pair $\varphi^{(k)}$, $\psi^{(k)}$ of V_k-valued functions, with $\varphi^{(k)} \in \mathcal{C}_0^\infty(G_k)$ supported in the unit ball, and $\psi^{(k)} \in \mathcal{S}(G_k)$ a Schwartz function, so that $\int_{G_k} \varphi^{(k)}(\mathbf{x}) \, d\mathbf{x} = \int_{G_k} \psi^{(k)}(\mathbf{x}) \, d\mathbf{x} = 0$, and

$$(8.17) \qquad \int_0^\infty \psi_a^{(k)}(\mathbf{xy}^{-1}) \cdot \varphi_a^{(k)}(\mathbf{y}) \frac{da}{a} = \delta_0.$$

Here $\varphi_a^{(k)}(\mathbf{x}) = a^{-Q_k - Q_{k+1} - \cdots - Q_n} \varphi^{(k)}(\delta_{a^{-1}}(\mathbf{x}))$, with a similar definition for $\psi_a^{(k)}(\mathbf{x})$. Also $\cdot$ denotes the inner product in V_k. See Theorem 1.61 in [3].

We define operators $P_a^{(k)}$ and $Q_a^{(k)}$, acting on functions on G_k, by setting $P_a^{(k)}(f) = f * \varphi_a^{(k)}$ and $Q_a^{(k)}(f) = f * \psi_a^{(k)}$. Note that (8.17) shows that

$$(8.18) \qquad \int_0^\infty P_a^{(k)} \cdot Q_a^{(k)} \frac{da}{a} = \text{Id}.$$

Next, define the square functions S_k and $S_k^\#$ by setting

$$S_k(f)(\mathbf{x}) = \Big(\int_0^\infty |P_a^{(k)}(f)(\mathbf{x})|^2 \frac{da}{a} \Big)^{1/2}, \quad S_k^\#(f)(\mathbf{x}) = \Big(\int_0^\infty |Q_a^{(k)}(f)(\mathbf{x})|^2 \frac{da}{a} \Big)^{1/2}.$$

The usual theory of singular integrals in [11] and [3], together with (8.18), then gives the inequalities

$$(8.19) \qquad \| f \|_{L^p} \approx \| S_k(f) \|_{L^p} \approx \| S_k^\#(f) \|_{L^p}$$

for $1 < p < \infty$ on G_k. The result is valid not only for scalar-valued f, but also for f that take their values in a Hilbert space.

Having recalled the known results for G_k we transfer them to the whole group G by writing

$$\tilde{P}_a^{(k)}(f) = f * (\delta_{\mathbf{x}_1\cdots\mathbf{x}_{k-1}} \otimes \varphi_a^{(k)}), \qquad \tilde{Q}_a^{(k)}(f) = f * (\delta_{\mathbf{x}_1\cdots\mathbf{x}_{k-1}} \otimes \psi_a^{(k)}),$$

$$\tilde{S}_k(f) = \left(\int_0^\infty |\tilde{P}_a^{(k)}(f)|^2 \frac{da}{a}\right)^{1/2}, \qquad \tilde{S}_k^{\#}(f) = \left(\int_0^\infty |\tilde{Q}_a^{(k)}(f)|^2 \frac{da}{a}\right)^{1/2}.$$

We get as a consequence

(8.18')
$$\int_0^\infty \tilde{P}_a^{(k)} \cdot \tilde{Q}_a^{(k)} \frac{da}{a} = \mathrm{Id}$$

on G, and

(8.19')
$$\| f \|_{L^p(G)} \approx \| \tilde{S}_k(f) \|_{L^p(G)} \approx \| \tilde{S}^{\#}(f) \|_{L^p(G)}.$$

With the above one-parameter theory arising from each G_k we come to the square functions on G that are relevant for us. For each $\mathbf{t} = (t_1, \ldots t_n) \in (\mathbb{R}^+)^n$, we set $P_\mathbf{t} = \tilde{P}_{t_n}^{(n)} \cdot \tilde{P}_{t_{n-1}}^{(n-1)} \cdots \tilde{P}_{t_1}^{(1)}$. That is, $P_\mathbf{t}(f) = f * \Phi_\mathbf{t}$, where $\Phi_\mathbf{t} = \tilde{\varphi}_{t_1}^{(1)} * \tilde{\varphi}_{t_2}^{(2)} * \cdots * \tilde{\varphi}_{t_n}^{(n)}$ and $\tilde{\varphi}_{t_k}^{(k)} = \delta_{\mathbf{x}_1\cdots\mathbf{x}_{k-1}} \otimes \varphi_{t_k}^{(k)}$. Note also that $\Phi_\mathbf{t}$ is a V-valued function, where $V = V_1 \otimes V_2 \otimes \cdots \otimes V_n$. Similarly, we define $P_\mathbf{t}^* = \tilde{P}_{t_1}^{(1)} \cdot \tilde{P}_{t_2}^{(2)} \cdots \tilde{P}_{t_n}^{(n)}$, $Q_\mathbf{t} = \tilde{Q}_{t_n}^{(n)} \cdots \tilde{Q}_{t_1}^{(1)}$, and $Q_\mathbf{t}^* = \tilde{Q}_{t_1}^{(1)} \cdots \tilde{Q}_{t_n}^{(n)}$. Also $Q_\mathbf{t}(f) = f * \bar{\psi}_\mathbf{t}$, with $\bar{\psi}_\mathbf{t} = \tilde{\psi}_{t_1}^{(1)} * \cdots * \tilde{\psi}_{t_n}^{(n)}$ and $\bar{\psi}_\mathbf{t}$ is also V-valued. Finally, we set

$$S(f)(\mathbf{x}) = \left(\int_{(\mathbb{R}^+)^n} |P_\mathbf{t}(f)|^2 \frac{d\mathbf{t}}{[\mathbf{t}]}\right)^{1/2}, \quad \mathfrak{S}(f)(\mathbf{x}) = \left(\int_{(\mathbb{R}^+)^n} |\mathcal{M}Q_\mathbf{t}f|^2 \frac{d\mathbf{t}}{[\mathbf{t}]}\right)^{1/2}.$$

Here we use the abbreviation that $[\mathbf{t}] = t_1 \cdot t_2 \cdots \cdot t_n$.

Lemma 8.12. *We have*

(a) $\displaystyle\int_{(\mathbb{R}^+)^n} P_\mathbf{t}^* Q_\mathbf{t} \frac{d\mathbf{t}}{[\mathbf{t}]} = \mathrm{Id};$

(b) $\| f \|_{L^p} \leq A_p \| S(f) \|_{L^p}$, $1 < p < \infty;$

(c) $\| \mathfrak{S}(f) \|_{L^p} \leq A_p \| f \|_{L^p}$, $1 < p < \infty.$

To prove (a), we take first the identity (8.18) when $k = 1$, and $a = t_1$. Next we multiply on the left of both sides by $P_{t_2}^{(2)}$ and on the right of both sides by $\tilde{Q}_{t_2}^{(2)}$ and integrate in t_2, using (8.18') for $k = 2$. Continuing this way yields (a). Inequality (b) follows from repeated comparisons of the corresponding inequalities (8.19') for $\tilde{S}_k$. Also (c) follows by applying (8.19') for $\tilde{S}_k^{\#}$ and a two-fold application of the vector-valued maximal function in Corollary 8.4, for $F_\mathbf{t}(\mathbf{x}) = Q_\mathbf{t}(f)(\mathbf{x}) \cdot [\mathbf{t}]^{-1/2}$.

The final lemma needed is as follows:

Lemma 8.13. *Suppose $\mathcal{K}$ is a flag kernel and $T(f) = f * \mathcal{K}$. Then,*

(8.20)
$$S(T(f))(\mathbf{x}) \leq c\, \mathfrak{S}(f)(\mathbf{x}).$$

Proof. Now $P_{\mathbf{t}}T(f)(\mathbf{x}) = \int_{(\mathbb{R}^+)^n} P_{\mathbf{t}} \cdot TP_{\mathbf{s}}^* Q_{\mathbf{s}}(f)(\mathbf{x}) \frac{d\mathbf{s}}{[\mathbf{s}]}$ by part (a) of the previous lemma. Hence Theorem 8.10, part (b) shows that

$$|P_{\mathbf{t}}T(f)(\mathbf{x})| \leq \int_{(\mathbb{R}^+)^n} \gamma(\mathbf{s}, \mathbf{t}) \, \mathcal{M}(Q_{\mathbf{s}}(f))(\mathbf{x}) \, \frac{d\mathbf{s}}{[\mathbf{s}]},$$

where $\gamma(\mathbf{s}, \mathbf{t}) = c \left(\prod_{k=1}^n \min\left(\frac{s_k}{t_k}, \frac{t_k}{s_k}\right) \right)^{\delta}$. Thus,

$$|P_{\mathbf{t}}T(f)(\mathbf{x})|^2 \leq \int_{(\mathbb{R}^+)^n} \gamma(\mathbf{s}, \mathbf{t})(\mathcal{M}Q_{\mathbf{s}}(f)(\mathbf{x}))^2 \, \frac{d\mathbf{s}}{[\mathbf{s}]} \cdot I(\mathbf{t})$$

with $I(\mathbf{t}) = \int_{(\mathbb{R}^+)^n} \gamma(\mathbf{s}, \mathbf{t}) \frac{d\mathbf{s}}{[\mathbf{s}]}$, by Schwarz's inequality. But $\sup_{\mathbf{t}} I(\mathbf{t}) = A < \infty$, since

$$\int_0^\infty \min\left(\frac{s_k}{t_k}, \frac{t_k}{s_k}\right)^{\delta} \frac{ds_k}{s_k} = t_k^{-\delta} \int_0^{t_k} s_k^{-1+\delta} \, ds_k + t_k^{\delta} \int_{t_k}^\infty s_k^{-1-\delta} \, ds_k = \frac{2}{\delta}.$$

A further integration in $\mathbf{t}$ (noting that also $\int_{(\mathbb{R}^+)^n} \gamma(\mathbf{s}, \mathbf{t}) \frac{d\mathbf{t}}{[\mathbf{t}]} \leq A$) then gives the desired result. $\square$

Theorem 8.14. *With $Tf = f * K$ as above, we have $\| Tf \|_{L^p(G)} \leq A_p \| f \|_{L^p(G)}$, for $1 < p < \infty$.*

This now follows directly from (8.20), once we apply Lemma 8.12, part (b), for Tf in place of f, and then part (c) of that Lemma.

9. Appendix I: Lifting

Suppose that $T(f) = f * \mathcal{K}$ is a convolution operator on G_k with $f \in \mathcal{S}(G_k)$ and $\mathcal{K} \in \mathcal{S}'(G_k)$, a tempered distribution. Then Tf can be written $T(f)(\mathbf{x}) = \int_{G_k} \mathcal{K}(\mathbf{y}) f(\mathbf{x}\mathbf{y}^{-1}) \, d\mathbf{y} = \langle \mathcal{K}, F_{\mathbf{x}} \rangle$, where $F_{\mathbf{x}}$ is the element of $\mathcal{S}(G_k)$ given by $F_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}\mathbf{y}^{-1})$ for $\mathbf{y} \in G_k$. We can lift T to a convolution operator on G, denoted by $\tilde{T}$, given by $\tilde{T}(f)(\mathbf{x}) = \int_{G_k} K(\mathbf{y}) f(\mathbf{x}\mathbf{y}^{-1}) \, d\mathbf{y} = \langle \mathcal{K}, F_{\mathbf{x}} \rangle$, where $f \in \mathcal{S}(G)$, and $F_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}\mathbf{y}^{-1})$ for $\mathbf{y} \in G_k$.

We describe this lifting in terms of the coordinate system used in the proof of Theorem 8.9. We can write each $\mathbf{x} \in G$ as a product $\mathbf{x} = \mathbf{x}' \cdot \bar{\mathbf{x}}$, with $\mathbf{x}' = (\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, 0, \ldots, 0)$ and $\bar{\mathbf{x}} \in G_k$, where $\bar{\mathbf{x}} = (0, \ldots, 0, \mathbf{x}_k, \ldots, \mathbf{x}_n)$. With this coordinate system, we define $\tilde{K} \in \mathcal{S}'(G)$ as $\delta_{\mathbf{x}'} \otimes K$, where $\mathbf{x} = \mathbf{x}' \cdot \bar{\mathbf{x}}$, and we set $\tilde{T}(f)(\mathbf{x}) = (f * \tilde{K})(\mathbf{x})$ for $f \in \mathcal{S}(G)$ and $\mathbf{x} \in G$. Then $\tilde{T}(f)(\mathbf{x}) = \tilde{T}(f)(\mathbf{x}' \cdot \bar{\mathbf{x}})$. However observe that $\tilde{T}(f)(\mathbf{x}' \cdot \bar{\mathbf{x}}) = T(f^{\mathbf{x}'})(\bar{\mathbf{x}})$, where $f^{\mathbf{x}'}$ is the element of $\mathcal{S}(G_k)$ given by $f^{\mathbf{x}'}(\mathbf{y}) = f(\mathbf{x}' \cdot \mathbf{y})$, $\mathbf{y} \in G_k$. Therefore

(9.1) $$\tilde{T}(f)(\mathbf{x}' \cdot \bar{\mathbf{x}}) = T(f^{\mathbf{x}'})(\bar{\mathbf{x}})$$

Next, suppose that T satisfies the bound

(9.2) $$\| T(f) \|_{L^p(G_k)} \leq A \| f \|_{L^p(G_k)}$$

for each $f \in \mathcal{S}(G)$. Then, applying this to $f = f^{\mathbf{x}'}$ via (9.1) (and assuming $p < \infty$), gives

$$\int_{G_k} |\tilde{T}(f)(\mathbf{x}'\bar{\mathbf{x}})|^p \, d\bar{\mathbf{x}} \leq A^p \int_{G_k} |f^{\mathbf{x}'}(\bar{\mathbf{x}})|^p \, d\bar{\mathbf{x}},$$

for each $\mathbf{x}'$, and an integration in $\mathbf{x}'$ yields

(9.3) $\| T(f) \|_{L^p(G)} \leq A \| f \|_{L^p(G)}$.

Suppose next that K depends on a parameter t, $K = K_t$ and set $T_t(f) = f * K_t$. Then the same argument shows that

$$\big\| \sup_t |\tilde{T}_t(f)| \big\|_{L^p(G)} \leq A \| f \|_{L^p(G)}, \quad \text{for all } f \in \mathcal{S}(G)$$

whenever

$$\big\| \sup_t |T_t(f)| \big\|_{L^p(G_k)} \leq A \| f \|_{L^p(G_k)}, \quad \text{for all } f \in \mathcal{S}(G_k).$$

This proves that the lifted maximal inequality $\| \tilde{M}_k(f) \|_{L^p(G)} \leq A_p \| f \|_{L^p(G)}$ follows from the corresponding inequality on G_k, by considering first the case when f is non-negative (and $K_t = \eta_{B_t}^{(k)}$), and then by replacing f by $|f|$.

In the same way the vector-valued maximal inequality (8.4) on G_k can be lifted to the corresponding inequality (8.4$'$) on G. In fact, it suffices to prove (8.4$'$) when there are only m non-zero f_j's, $1 \leq j \leq m$, with bounds independent of m. With this understanding, set

$$f = (f_1, \ldots, f_m), \qquad t = (t_1, \ldots, t_m),$$

and

$$T_t(f) = (T_{t_1}^1(f_1), \ldots, T_{t_m}^m(f_m)), \qquad T_{t_j}^j(f) = f * K_{t_j}^j,$$

with

$$|f| = \Big(\sum_{j=1}^m |f_j|^2 \Big)^{1/2}, \qquad |T_t(f)| = \Big(\sum_{j=1}^m |T_{t_j}^j(f_j)|^2 \Big)^{1/2}.$$

Note that

$$\Big(\sum_{j=1}^m \sup_{t_j} |T_{t_j}^j(f)|^2 \Big)^{1/2} = \sup_t |T_t|(f)|.$$

Then as before the inequality

$$\big\| \sup_t |T_t(f)| \big\|_{L^p(G_n)} \leq A \| |f| \|_{L^p(G_k)}, \quad f \in \mathcal{S}(G_n)$$

implies the corresponding inequality for T_t lifted to G, that is for $\tilde{T}_t$, and this then yields the desired result.

We should remark that the lifting procedure used here can be viewed in terms of the more abstract "transference" method presented in [1].

10. Appendix II: An estimate for a geometric sum

Recall that $\mathcal{E}_n = \{I = (i_1, \ldots, i_n) \in \mathbb{Z}^n \,\big|\, i_1 \leq i_2 \leq \cdots \leq i_n\}$. More generally, if $0 \leq B_1 \leq B_2 \leq \cdots \leq B_n$, let

$$(10.1) \qquad \Lambda(B) = \left\{I = (i_1, \ldots, i_n) \in \mathcal{E}_n \,\big|\, B_j \leq 2^{i_j} \quad \text{for} \quad 1 \leq j \leq n\right\}.$$

We establish the following estimate for geometric sums which we shall use several times.

Proposition 10.1. *Let $\alpha_j > 0$ for $1 \leq j \leq n$, and let $M > \sum_{k=1}^n \alpha_k$. There is a constant C depending on n, on M, and on the numbers $\{\alpha_j\}$ so that for any $A_1, \ldots, A_n \in (0, \infty)$ and any $0 \leq B_1 \leq B_2 \leq \cdots \leq B_n$,*

$$(10.2) \qquad \sum_{I \in \mathcal{E}_n} \frac{\prod_{k=1}^n (2^{-i_k})^{\alpha_k}}{\left(1 + \sum_{k=1}^n 2^{-i_k} A_k\right)^M} \leq C \prod_{j=1}^n (A_1 + A_2 + \cdots + A_j)^{-\alpha_j},$$

$$(10.3) \qquad \sum_{I \in \Lambda(B)} \frac{\prod_{k=1}^n (2^{-i_k})^{\alpha_k}}{\left(1 + \sum_{k=1}^n 2^{-i_k} A_k\right)^M} \leq C \prod_{j=1}^n (A_1 + A_2 + \cdots + A_j + B_j)^{-\alpha_j}.$$

Proof. Note that if we take $B_1 = \cdots = B_n = 0$, then the inequality (10.3) gives the inequality (10.2). If for $x_1, \ldots, x_n \geq 0$ we put

$$\varphi(x_1, \ldots, x_n) = \left(\prod_{k=1}^n x_k^{\alpha_k}\right)\left(1 + \sum_{k=1}^n x_k A_k\right)^{-M},$$

then, if $1 \leq s_j \leq 2$ for $1 \leq j \leq n$, we have

$$\varphi(s_1 x_1, \ldots, s_n x_n) \leq 2^{|\alpha|} \prod_{k=1}^n x_k^{\alpha_k} \left(1 + \sum_{k=1}^n x_k A_k\right)^{-M} \leq 2^{|\alpha|} \varphi(x_1, \ldots, x_n)$$

For each $I = (i_1, \ldots, i_n) \in \mathbb{Z}^n$, let $Q_I = \{\mathbf{x} \in \mathbb{R}^n \,\big|\, \frac{1}{2} 2^{-i_k} \leq x_k < 2^{-i_k}\}$. It follows that there is a constant $C = C(n, \alpha)$ depending only on n and α such that

$$C^{-1} \int_{Q_I} \varphi(\mathbf{x}) \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n} \leq \frac{\prod_{k=1}^n (2^{-i_k})^{\alpha_k}}{\left(1 + \sum_{k=1}^n 2^{-i_k} A_k\right)^M} \leq C \int_{Q_I} \varphi(\mathbf{x}) \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n}.$$

Thus

$$\sum_{I \in \Lambda(B)} \frac{\prod_{k=1}^n (2^{-i_k})^{\alpha_k}}{\left(1 + \sum_{k=1}^n 2^{-i_k} A_k\right)^M} \leq C \int_{\bigcup_{I \in \Lambda(B)} Q_I} \varphi(\mathbf{x}) \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n}.$$

On the other hand, it is easy to check that if $I \in \Lambda(B)$ and $\mathbf{x} \in Q_I$, then $0 < x_{k+1} \leq 2 x_k$ for $1 \leq k \leq n-1$ and $x_k \leq B_k^{-1}$ for $1 \leq k \leq n$. Thus if we put

$$\Omega(B) = \left\{\mathbf{x} \in \mathbb{R}_+^n \; : \; \begin{cases} \frac{1}{2} x_{k+1} \leq x_k \leq B_k^{-1} & \text{for } 1 \leq k \leq n-1 \\ 0 \leq x_n \leq B_n^{-1} \end{cases}\right\},$$

then $\bigcup_{I \in \Lambda(B)} Q_I \subseteq \Omega(B)$. Thus to prove the proposition, it suffices to show that

$$\int_{\Omega(B)} \frac{\prod_{k=1}^n x_k^{\alpha_k}}{\left(1 + \sum_{k=1}^n x_k A_k\right)^M} \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n} \leq C \prod_{j=1}^n (A_1 + A_2 + \cdots + A_j + B_j)^{-\alpha_j}.$$

However,

$$\int_{\Omega(B)} \prod_{k=1}^n x_k^{\alpha_k} \left(1 + \sum_{k=1}^n x_k A_k\right)^{-M} \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n}$$

$$= \int_0^\infty t^M e^{-t} \left[\int_{\Omega(B)} \prod_{k=1}^n x_k^{\alpha_k} e^{-x_k A_k t} \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n} \right] \frac{dt}{t}.$$

We will show that we can estimate the inner integral on the right hand side by

(10.4)
$$\int_{\Omega(B)} \prod_{k=1}^n x_k^{\alpha_k} e^{-x_k A_k t} \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n}$$

$$\leq C(n, \alpha) \prod_{j=1}^n (1 + t^{-\alpha_j})(A_1 + A_2 + \cdots + A_j + B_j)^{-\alpha_j},$$

and since $M > \sum_{k=1}^n \alpha_k$, this will complete the proof.

To establish (10.4), we first establish an estimate for $\int_{\frac{x}{2}}^{B^{-1}} s^\alpha e^{-sAt} \frac{ds}{s}$. On the one hand, we have

$$\int_{\frac{x}{2}}^{B^{-1}} s^\alpha e^{-sAt} \frac{ds}{s} \leq e^{-\frac{1}{2}xAt} \int_0^{B^{-1}} s^{\alpha-1} ds = \alpha^{-1} e^{-\frac{1}{2}xAt} B^{-\alpha}.$$

On the other hand we have

$$\int_{\frac{x}{2}}^{B^{-1}} s^\alpha e^{-sAt} \frac{ds}{s} \leq (At)^{-\alpha} \int_{\frac{1}{2}xAt}^\infty s^{\alpha-1} e^{-s} ds \leq C_\alpha (At)^{-\alpha} e^{-\frac{1}{4}xAt}.$$

Putting the two together, we have the estimate

(10.5) $$\int_{\frac{x}{2}}^{B^{-1}} s^\alpha e^{-sAt} \frac{ds}{s} \leq C_\alpha e^{-\frac{1}{4}xAt}(At + B)^{-\alpha} \leq C_\alpha e^{-\frac{1}{4}xAt}(A + B)^{-\alpha}(1 + t^{-\alpha}).$$

We now establish (10.4) by induction on n. When $n = 1$, we use (10.5) with $x = 0$ to get

$$\int_0^{B_1^{-1}} x_1^{\alpha_1} e^{-\alpha_1 A_1 t} \frac{dx_1}{x_1} \leq C_\alpha (A_1 + B_1)^{-\alpha_1}(1 + t^{-\alpha_1}).$$

For the induction step, we have

$$
\int_{\Omega(B)} \prod_{k=1}^{n} x_k^{\alpha_k} \, e^{-x_k A_k t} \, \frac{dx_1 \cdots dx_n}{x_1 \cdots x_n}
$$

$$
= \int_{\Omega'(B)} \prod_{k=2}^{n} x_k^{\alpha_k} e^{-x_k A_k t} \Big[\int_{\frac{x_2}{2}}^{B_1^{-1}} x_1^{\alpha_1} e^{-\alpha_1 A_1 t} \, \frac{dx_1}{x_1} \Big] \frac{dx_2 \cdots dx_n}{x_2 \cdots x_n}
$$

$$
\leq C_\alpha (A_1 + B_1)^{-\alpha_1} (1 + t^{-\alpha_1}) \int_{\Omega'(B)} \prod_{k=2}^{n} x_k^{\alpha_k} e^{-x_k A_k t} e^{-\frac{1}{4} x_2 A_1 t} \frac{dx_2 \cdots dx_n}{x_2 \cdots x_n}.
$$

Here,

$$
\Omega'(B) = \left\{ \mathbf{x} \in \mathbb{R}_+^n \ : \ \left\{ \begin{array}{ll} \frac{1}{2} x_{k+1} \leq x_k \leq B_k^{-1} & \text{for } 2 \leq k \leq n-1 \\ 0 \leq x_n \leq B_n^{-1} \end{array} \right. \right\},
$$

and we have used the estimate in (10.5). The last integral on the right-hand side is thus of the same form as the original integral, except that n has been replaced by $n-1$, and A_2 has been replaced by $A_2 + \frac{1}{4} A_1$. We can thus use our inductive hypothesis on this integral, and we obtain the desired estimate. This completes the proof.   □

References

[1] Coifman, R. R, and Weiss, G.: *Transference methods in Analysis.* CBMS Reg. Conf. Series in Math., 31. American Mathematical Society, 1976.

[2] Ding, Y., Lu, G. Z. and Ma, B. L.: Multi-parameter Triebel–Lizorkin and Besov spaces associated with flag singular integrals. *Acta Math. Sin. (Engl. Ser.)* **26** (2010), no. 4, 603–620.

[3] Folland, G. B. and Stein, E. M.: *Hardy spaces on homogeneous groups.* Mathematical Notes, 28. Princeton University Press, Princeton, NJ, 1982.

[4] Głowacki, P.: Composition and L^2-boundedness of flag kernels. *Colloq. Math.* **118** (2010), no. 2, 581–585.

[5] Głowacki, P.: Correction to "Composition and L^2-boundedness of flag kernels". *Colloq. Math.* **120** (2010), no. 2, 331.

[6] Głowacki, P.: L^p-boundedness of flag kernels on homogeneous groups. Preprint.

[7] Müller, D., Ricci, F. and Stein, E. M.: Marcinkiewicz multipliers and multi-parameter structure on Heisenberg (-type) groups. I. *Invent. Math.* **119** (1995), 119–233.

[8] Nagel, A., Ricci, F. and Stein, E. M.: Harmonic analysis and fundamental solutions on nilpotent Lie groups. In *Analysis and partial differential equations,* 249–275. Lecture Notes in Pure and Appl. Math., 122. Dekker, New York, 1990.

[9] Nagel, A., Ricci, F. and Stein, E. M.: Singular integrals with flag kernels and analysis on quadratic CR manifolds. *J. Functional Analysis* **181** (2001), 29–118.

[10] Nagel, A., Ricci, F., Stein, E. M. and Wainger, S.: Singular integrals with flag kernels on homogeneous groups, II. To appear.

[11] Stein, E. M.: *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals.* Princeton Mathematical Series, 43. Princeton University Press, Princeton, New Jersey, 1993.

[12] Street, B.: Multi-parameter singular radon transforms I: the L^2 theory. *J. Anal. Math.* **116** (2012), no. 1, 83–162.

[13] Yang, D.: Besov and Triebel–Lizorkin spaces related to singular integrals with flag kernels. *Rev. Mat. Complut.* **22** (2009), no. 1, 253–302.

Alexander Nagel: Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA.
E-mail: nagel@math.wisc.edu

Fulvio Ricci: Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy.
E-mail: fricci@sns.it

Elias M. Stein: Department of Mathematics, Princeton University, Princeton, NJ 08544, USA.
E-mail: stein@math.princeton.edu

Stephen Wainger: Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA.
E-mail: wainger@math.wisc.edu